OECD
Programme for International Student Assessment:
Unraveling a Knowledge Network

by

Clara Morgan

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

School of Public Policy and Administration
Carleton University
Ottawa, Ontario
September 2007
© Copyright
2007, Clara Morgan
NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.
Abstract

The dissertation examines the construction of an international student assessment, the Programme for International Student Assessment (PISA), by the Organization for International Co-operation and Development (OECD). The PISA is analyzed within the context of the global architecture of education where various agents such as the OECD and other international organizations, states and communities of experts are reconstructing, reproducing and legitimizing the discourse and material practices of societal and economic progress and of official educational knowledge. The dissertation argues that the PISA is a fragile entity that is susceptible to contestation not only as it is located in a highly politicized setting, the OECD, but also because it is founded on the socially-constructed science of educational measurement.

The dissertation draws on the Foucauldian concept of a power bloc formation in order to analyze the PISA’s technical capacity, relations of communication and relationships of power. The analysis contributes to our understanding of how programs such as the PISA come to be constructed, how they work, how they attempt to order and shape our worlds, and how they are connected to global networks of power. It reveals that the international educational statistics that inform public policy decision-making are the result of a series of negotiated compromises made by communities of practice occupying various institutional structures within the global architecture of education.
Acknowledgements

This work would not have been possible without the kind and generous support of several individuals and institutions. I would like to thank my dissertation committee, Bruce Curtis, Rianne Mahon, and Lisa Mills, for their guidance, support, optimism and valuable advice during my intellectual journey.

I am grateful for the support provided by the students, staff and faculty of the School of Public Policy and Administration and for the financial support I received in the form of two Ontario Graduate Scholarships and scholarships from Carleton University.

I would like to thank the interview respondents who participated in my study and who generously gave from their time.

Four people deserve a special and loving thank you: my husband, Adam, and my children, Emilia, Morjan and Alicia. Without their unwavering support, I could not have survived this journey.
# Table of Contents

<table>
<thead>
<tr>
<th>Acceptance Form</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables and Figures</td>
<td>vii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>ix</td>
</tr>
</tbody>
</table>

Chapter 1
Introduction.............................................1

Chapter 2
Concepts, Ideas and Methodology.................5

Chapter 3
Building Technical Capacity in Educational Measurement..37

Chapter 4
The International Association for the Evaluation of Educational Achievement.......59

Chapter 5
Educational Activities at the OECD................89

Chapter 6
American Influence and the Adjustment of the Power Bloc..................112

Chapter 7
The Formation of the OECD PISA.......................141

Chapter 8
Power Relations........................................176

Chapter 9
The Global Architecture of Education and the Fragility of the PISA.............215

Chapter 10
Conclusion.............................................240

Appendix 1 List of Interviewees......................251
Appendix 2 Consent Form...............................252
Appendix 3 List of Questions.........................254
References
Published Books, Reports and Articles 255
Other References 280
List of Tables and Figures

TABLES

Table 1: Political Economic Context and Education Policy Developments 18
Table 2: The Six Subject Study 76
Table 3: List of IEA Studies 87
Table 4: Education inside the OECD – Committees and Governing Boards 100
Table 5: OECD Educational Statistics and Indicators – Chronology of Infrastructural Developments 137
Table 6: OECD PISA Developments 160
Table 7: The Literacy Domains in the IALS 179
Table 8: Literacy Scale for Prose, Document and Quantitative 179
Table 9: IALS Literacy Abilities 179
Table 10: Measuring PISA Reading Abilities 183
Table 11: Measuring PISA Mathematics Abilities 186
Table 12: Measuring PISA Scientific Abilities 186

FIGURES

Figure 1: A hierarchical factor model of ‘g’ 40
Figure 2: An example of an Item Characteristic Curve 54
Figure 3: ISCED Classification of Educational Programs (1976) 109
Figure 4: ISCED Classification of Educational Programs (1997) 130
Figure 5: An example of an immutable mobile: the Student Tracking Form 174

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 6</td>
<td>Education at the OECD – Committee &amp; Governing Board Structure</td>
<td>193</td>
</tr>
<tr>
<td>Figure 7</td>
<td>The Relationship of the BPC to Network A (1997)</td>
<td>195</td>
</tr>
<tr>
<td>Figure 8</td>
<td>The Relationship of Network A to PGB (present)</td>
<td>196</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Organizational Components of the OECD PISA</td>
<td>197</td>
</tr>
<tr>
<td>Figure 10</td>
<td>An example from an OECD PISA publication reporting PISA 2000 results</td>
<td>217</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Sample of student performance rankings with American populations disaggregated by ethnic grouping</td>
<td>220</td>
</tr>
</tbody>
</table>
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACER</td>
<td>Australian Council for Educational Research</td>
</tr>
<tr>
<td>ANT</td>
<td>Actor-Network Theory</td>
</tr>
<tr>
<td>BISCE</td>
<td>Board on International Comparative Studies in Education</td>
</tr>
<tr>
<td>BPC</td>
<td>Board of Participating Countries</td>
</tr>
<tr>
<td>CTT</td>
<td>Classical Test Theory</td>
</tr>
<tr>
<td>CERI</td>
<td>Centre for Educational Research and Innovation</td>
</tr>
<tr>
<td>Cito</td>
<td>Netherlands National Institute for Educational Measurement</td>
</tr>
<tr>
<td>CSTP</td>
<td>Committee for Scientific and Technical Personnel</td>
</tr>
<tr>
<td>ETS</td>
<td>Educational Testing Service</td>
</tr>
<tr>
<td>IAEP</td>
<td>International Assessment of Educational Progress</td>
</tr>
<tr>
<td>IALS</td>
<td>International Adult Literacy Survey</td>
</tr>
<tr>
<td>IEA</td>
<td>International Association for the Evaluation of Educational Achievement</td>
</tr>
<tr>
<td>IMHE</td>
<td>Programme on Institutional Management in Higher Education</td>
</tr>
<tr>
<td>INES Project</td>
<td>International Indicators of Education Systems Project</td>
</tr>
<tr>
<td>IRT</td>
<td>Item Response Theory</td>
</tr>
<tr>
<td>ISCED</td>
<td>International Standard Classification of Education</td>
</tr>
<tr>
<td>NAEP</td>
<td>National Assessment of Educational Progress</td>
</tr>
<tr>
<td>NCES</td>
<td>National Center for Education Statistics</td>
</tr>
<tr>
<td>NDEA</td>
<td>National Defense Education Act</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>NFER</td>
<td>National Foundation for Educational Research in England and Wales</td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Development and Cooperation</td>
</tr>
<tr>
<td>OEEC</td>
<td>Organisation for European Economic Cooperation</td>
</tr>
<tr>
<td>OSTP</td>
<td>Office for Scientific and Technical Personnel</td>
</tr>
<tr>
<td>PEB</td>
<td>Programme on Educational Building</td>
</tr>
<tr>
<td>PGB</td>
<td>PISA Governing Board</td>
</tr>
<tr>
<td>PISA</td>
<td>Programme for International Student Assessment</td>
</tr>
<tr>
<td>TAG</td>
<td>Technical Advisory Group</td>
</tr>
<tr>
<td>TIMSS</td>
<td>Trends in Mathematics and Science Study</td>
</tr>
<tr>
<td>UIE</td>
<td>UNESCO Institute of Education</td>
</tr>
<tr>
<td>UNESCO</td>
<td>United Nations Educational, Scientific and Cultural Organization</td>
</tr>
<tr>
<td>YALS</td>
<td>Young Adult Literacy Survey</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

My dissertation examines the construction of an international student assessment, the Programme for International Student Assessment (PISA), within the “global architecture of education” (Jones 2006: 43).¹ The PISA is administered by the Organization for Economic Co-operation and Development (OECD), an international governmental organization of elite industrialized states (Salzman and Terracino 2006). In this global architecture of education, various agents such as the OECD and other international organizations, states and communities of experts are reconstructing, reproducing and legitimizing the discourse and material practices of societal and economic progress (Chabbott 2003) and of official educational knowledge² (Apple 1996, 2000; Apple and Oliver 2003).

The dissertation analyzes the PISA as a process established by OECD member states to measure the quality of their school systems and to monitor levels of student achievement using standardized achievement tests and survey questionnaires. The PISA is analyzed as a fragile entity that is susceptible to contestation not only as it is located in a highly politicized setting, the OECD, but

¹ By global architecture of education, I follow Jones (2006) who describes it as a “system of global power relations that exerts a heavy, indeed determining, influence on how education is constructed around the world” (43). In the chapters that follow, I will describe in greater detail the global architecture of education and the role that both the OECD and the PISA play in it.
² I borrow the term “official knowledge” from Michael Apple who has used it in his writings to refer to common-sense or legitimated knowledge that is promulgated by an alliance forged among neoconservative and neoliberal thinkers and professionals (see especially 2000: xxiv-xxviii).
also because it is founded on the socially-constructed science of educational measurement.

In tracing the construction of the PISA, the dissertation seeks to answer the following research questions: (1) How did an infrastructure for international student assessment come to be constructed under the auspices of an international economic organization? (2) What were the conditions that led to its construction? (3) How does the PISA, as part of the global architecture of education, help to reinforce dominant educational discourse, values and material practices? In answering these questions, the dissertation reveals how actors and agents actively shape and define official educational knowledge.

The dissertation draws on the Foucauldian concept of a power bloc formation in order to analyze the OECD PISA. The OECD PISA is described as a power bloc formation that exercises power as it builds its technical capacity, relations of communication and power relations. The “adjustment of abilities, the resources of communication, and power relations constitute regulated and concerted systems” (Foucault 1983: 218). The dissertation examines the PISA’s technical capacity by tracing the history of international student assessments and analyzing the tools used by community members in developing technical expertise. It looks at PISA’s relations of communication by studying how work is coordinated and negotiated across communities of practice. The PISA’s relationships of power are analyzed by examining how the PISA governs the

---

3 The concept of the power bloc will be explained in greater detail in the next chapter.
conduct of individuals and how knowledge production is controlled and legitimated through institutionalized structures.

The purpose of the dissertation is to contribute to our understanding of how programs such as the PISA come to be constructed, how they work, how they attempt to order and shape our worlds, and how they are connected to global networks of power. As the dissertation demonstrates, the PISA is an integral component in the global architecture of education, connecting communities of practitioners across levels of governance and across international organizational structures. The PISA becomes one node in a global network that constructs official knowledge and material practices.

In the dissertation, I argue that it is not merely the ideational practices that need to concern us as researchers in the public policy field. We need also to be concerned with analyzing the material practices and relationships that interact and overlap to form a power bloc for exercising power. The power bloc formation of the PISA remains fragile. Despite the political and scientific work that has informed the PISA's construction, it remains vulnerable. The controversies that surrounded the field of comparative international education continue to haunt the PISA. The PISA's institutional location within the OECD weakens it, as it remains susceptible to political conflicts that arise among member states and to the power struggles among various agents. The dissertation reveals that the international educational statistics that inform public policy decision-making are the result of a series of negotiated compromises.
made by communities of practice occupying various institutional structures within the global architecture of education.
Chapter 2
Concepts, Ideas and Methods

Introduction
My research explores the history of the emergence, creation, and institutionalization of an international assessment called the Programme for International Student Assessment (PISA) that was formed under the auspices of the Organization for Economic Cooperation and Development (OECD). My purpose is to map the growth of international student assessments using the PISA as my case study. I seek to understand the role that international organizations such as the OECD play in constructing international student assessments. Such assessments produce indicators that serve to address public policy objectives and bureaucratic ends and to sustain a certain political economic rationality. The process of constructing indicators involves attempts at standardizing, classifying and categorizing objects to fit them into the representational schemes that make up our worlds.

This chapter explains my theoretical approach, methods and tools. I begin by explaining what the OECD PISA is as it is described in the OECD discourse. The next section focuses on my theoretical framework. It provides an overview of the concepts and theories that I use in my research work. I then discuss the manner in which I will study the OECD PISA, outlining the contents of each chapter. In the final section, I discuss the data gathering methods I used for my empirical research.

\[^{4}\text{I use the term 'OECD PISA' when speaking about the organizational actor. I use 'PISA' when referring to the assessment instrument.}\]
What is the OECD PISA? The Meta-Narrative Perspective

The Programme for International Student Assessment (PISA) assesses the reading, mathematics and science literacy and problem solving abilities of fifteen year olds in order to measure "how well young adults ... are prepared to meet the challenges of today's knowledge societies" (OECD 2001: 14). It collects background information on students, schools and principals by administering questionnaires. The first PISA was implemented in 2000 and subsequent assessments have been administered triennially. In its publications, the OECD defines the PISA as a measure of the "degree of preparedness of young people for adult life" (OECD 1999a: 11). Its indicators are deemed to be useful tools for comparing national educational systems and for answering the question, "how good is the education system itself?" (OECD 1996b: 24).

According to the OECD, the PISA was developed in response to the need for "solid and internationally comparative evidence of educational outcomes ... [by] ... governments and the general public" (1999a: 7). International educational assessments have existed prior to the creation of the OECD PISA, such as those delivered by the International Association for the Evaluation of Educational Achievement (IEA) and by the Education Testing Service's International Assessment of Educational Progress (IAEP) (OECD 1996; 1999). The OECD PISA distinguishes itself from such international assessments (a) by its origins, since it was initiated by OECD member countries; (b) by its regularity, since the PISA is administered every three years; (c) by its focus on assessing
fifteen year olds; and, (d) by its unique approach which tests literacy and skills rather than national curricula (OECD 1999:11).

As one of my interview respondents pointed out to me, “the PISA is a political success, not a scientific one” (Interview Respondent #8, 19 June 2006). The political ‘success’ of the OECD’s Programme for International Student Assessment (PISA) is shown by the growing number of countries participating in the study. Thirty-two states participated in the first PISA which took place in 2000. Participation grew dramatically from forty-one states in 2003 to fifty-seven in 2006. Of the fifty-seven states, twenty-seven did not belong to the OECD. Another measure of ‘success’ is the media attention the OECD PISA results draw and the public interest they arouse. These results are often reported using league tables and rankings comparing levels of achievement of student populations. This form of media reporting depicts school performance as an indicator of global competitiveness where students are ‘racing’ each other in a global ‘horse race’. The ‘winners’ are those students who score highest on the test (see for example, Repo 2005; Stack 2006).

In contrast to most OECD policy recommendations and best practices, which have little short-term impact on national political agendas, the OECD PISA results have a distinctive effect – one that causes such results to “enter national debates”, particularly for those countries which rank low or lower than expected on the PISA league tables (Armingeon 2004: 237). For example, the

---

5 In 2002, an additional 11 countries implemented the PISA 2000 assessment. The results are interpreted in terms of the PISA 2000 assessment cycle.
6 A list of participating countries is found on the PISA website for each PISA cycle. See: http://www.pisa.oecd.org/pages/0.2966.en_32252351_32236225_1 1 1 1 1.00.html, last accessed November 22, 2006.
German students' low ranking in the PISA 2000 spurred a nation-wide debate on educational policy and led to a series of educational reforms (Finetti 2004). Spokespeople of the OECD PISA point often to the German example to demonstrate the significance and the 'success' of this assessment. Analysis of media coverage of the OECD PISA 2000 in Canada shows that government press releases were used to inform the reporting, with little analysis of the results. In Canada, OECD PISA results were used to justify current provincial government education policies without delving into other contradictory data that showed, for example, low student interest or lack of parental support or achievement based on socio-economic status (Stack 2006).

My literature review on the OECD PISA turned up a number of studies analyzing the PISA results, including those conducted by the OECD, participating countries, educational researchers and those involved with the OECD PISA.7 The OECD has traditionally published reviews of national policies on education but has recently initiated a new analytical approach to these reviews that integrates the PISA results with expert analysis of educational systems. Other articles published in academic journals were written by

---

individuals involved with the OECD PISA. A few studies have used
disaggregated PISA results to show that a relationship exists between poverty
and student achievement (Berliner 2005; Lemke et al 2004). Despite such
voluminous knowledge production, there have been very few critical, analytical
studies of the OECD PISA.

What is the PISA? Splintering the Meta-Narrative

The OECD's discursive practices serve as a terrain for the OECD PISA to
propagate its official knowledge production within a particular political economic
context. What is being described by the meta-narrative of the OECD
publications, national reports and other secondary research is the end result of a
complex knowledge production process that generates international educational
indicators and reorders educational policy ideas and practices. In my
examination of the PISA, I will consider the literature that focuses on the
production of knowledge by international organizations and the
supranationalization of social and education policy.

Recent research in the public policy field has illustrated the important role
played by international organizations in transmitting and constructing knowledge
(see for example Sahlin-Andersson 2000; Porter and Webb 2004; Mahon 2005).

8 For example, Wynne Harlen, who was the Chair of the PISA Science Expert Group, has written
on the PISA's approach to assessing scientific literacy (see Harlen 2003). Expert members from
the Nordic countries, Pirjo Linnakylä and Svein Lie, have conducted secondary analyses of the
PISA results and have edited a report which analyzes PISA results from Denmark, Finland,
Norway or Sweden (see Lie, Linnakylä and Roe 2003).
9 Balzer and Bremen (2004) have examined how international organizations such as the OECD
influence national educational policy.
10 I have borrowed the term "splintering the metanarrative" from Slaughter (2001:391).
International organizations such as the OECD provide policy advice, prescriptions, and ideas. The knowledge they produce becomes a "guide to future directions in the reproduction and development of practices that shape an increasingly harmonized global political and economic system" (Porter and Webb 2004:1). International Relations researchers have analyzed the supranationalization of policy and pointed to the increased trend towards regulation, redistribution and provision occurring at global levels of governance (Deacon 1997). More recently, they have examined the reordering of the world through processes of transnational regulation (Djelic and Sahlin-Andersson 2006; Jacobsson and Sahlin-Andersson 2006).

My research continues with this research agenda by analyzing and documenting in greater detail the material practices and processes involved in the supranationalization of policy and of transnational governance. By taking the OECD PISA as my case study, I am able to analyze in a systematic manner the ways in which tools, processes and experts come together to forge an infrastructure that builds and expands the global architecture of education.

A seminal work that helps inform this study's understanding of OECD educational policy is that of Henry et al (2001), The OECD, Globalisation and Education Policy. Henry et al (2001) critically analyze how the OECD influences educational policy-making and the effects of globalization on national and international educational policy. Their study reveals that educational policy directions are continuously contested in a complex organization such as the OECD. Such contestation reflects "more ambiguous stances, which may
contribute to both strengthening and to undermining national policy making" (4).

Henry et al point to the ideological tensions that exist in OECD social democratic and neo-liberal policy approaches, often referred to by OECD insiders as ‘European’ versus ‘Anglo-Saxon’ (63). These tensions are manifested as the OECD tries to reconcile member states’ concerns for educational equity with educational quality, efficiency and effectiveness. Tensions also arise as OECD member states struggle to contain the negative repercussions of neoliberal policy approaches (such as increased income disparities, rising poverty rates, and marginalization). These solutions have been framed as part of the social inclusion/exclusion discourse (80-82).

Although Henry et al (2001) do not address the OECD PISA since their study was conducted prior to its creation, they do analyze the OECD’s educational indicators. They argue that OECD work on educational indicators has shaped national educational policies and priorities. Educational indicators instill a “corrective tendency” among OECD member states that “thrusts forward a degree of policy homogenisation” (102). OECD influence in this policy sphere has been achieved as the OECD took on a dual role – that of “policy instrument and forum – serving as a catalyst to facilitate policy development in member countries, assisting processes of policy dissemination, adaptation and borrowing – to assuming the status of an international mediator of knowledge and global policy actor” (84).

Henry et al view the OECD’s work on educational indicators as a mechanism for “merchandising a particular thesis of educational purposes (new
human capital theory) and of governance (new public management theory).”
They note that these “two items exist in tandem: new public management may
be regarded as the organisational template of the neo-liberal state” (103). The
OECD PISA provides me with a lens to observe the ways in which experts
transmit dominant theoretical frameworks, such as human capital and new
public management, through the associations they make, the tools they
construct and the work activities they practice.11

Theoretical Framework
The dissertation analyzes the ‘merchandising’ of a particular thesis of
educational purposes and the changes in educational policy and measurement
practices. It does so by drawing on the disciplinary traditions of political
economy, international relations, and sociology. Theories from political
economy situate the emergence of the OECD PISA within the larger context of
the neoliberal project that is constantly “mutating”. As Larner points out, the
coherence of such a project is achieved over time and needs to be understood
as “involving innovation, experimentation, and contestation” (Larner 2005: 2).
International political economic theories also situate the OECD within a large
network of international organizations that I discuss in terms of the global
architecture of education. I also draw on international relations theorists to help
explain the role of American actors within this global structure. Sociological
theories inform the dissertation’s analysis of power and organizational

11 Human capital theory will be discussed in greater detail in Chapter 3 and the New Public
Management Theory in Chapter 4.
processes. I use Michel Foucault's (1983) work to frame my analysis of the exercise of power and draw on his conceptualization of power as a formation of a "bloc of capacity-communication-power" (218). I integrate this Foucauldian understanding of power with concepts from the sociology of science and technology in order to analyze the role of the OECD in constructing the Programme for International Student Assessment (PISA).

Neoliberalism and the Neoliberal Project

The OECD's role in the construction of an international student achievement assessment needs to be understood within the broader context of neoliberalism. Neoliberalism\textsuperscript{12} promotes the market instead of the state as the regulator of the population and creates the entrepreneurial, active subject. Neoliberal policies minimize government intervention in the operation of the markets, conceive individuals as "active in making choices" and promote the exercise of choice by free individuals who are viewed as partners with the state (Rose 1999:142-143).

Neoliberalism began to take hold in the late 1970s and early 1980s as OECD member states turned away from Keynesian policies that were instituted in the post-war era (Harvey 2005: 39). With Margaret Thatcher in power in 1979 and Ronald Reagan in 1981, the public sector was to be drastically reshaped (Brown 2003). Neoliberal policy reforms were characterized by the retrenchment of the welfare state, privatization of the government sector, and the deregulation of state industries. At the same time, the degree to which neoliberal reforms took hold in OECD member states varied. Researchers have

\textsuperscript{12} Neoliberalism originated in the writings of intellectuals and academics such as Friedrich Hayek, Milton Friedman and various economists from the Chicago School (see Rose 1996, 1999; Brown 2003).
argued that the "depth and character of welfare state restructuring" was influenced by the "national configurations of democratic institutions" (Swank 2001: 198). The expansion of the welfare state in the post-war era was also accompanied by the "expansion of its constituencies" and hence of organized opposition to neoliberal reforms (Pierson 2001: 411). Even Thatcher, despite her aggressive privatization policies, was not able to privatize "core public services such as the national health care system and public education" (Harvey 2005: 88).

In the 1980s and 1990s, as part of the neoliberal reform agenda, governments adopted the New Public Management (NPM) theory\(^{13}\) and implemented private sector managerial techniques in the public policy sphere. Comparative studies indicate variation in the types of NPM reforms governments chose to adopt (Pal 2005; Pollitt and Summa 1997). Such variations may be attributable to macroeconomic conditions, political parties in power, administrative cultures, and national political systems. In the Scandinavian countries, for example, the implementation of NPM reforms differed because of party politics, the way in which political issues were framed, and which political party happened to be in power (Green-Pedersen 2002). Studies of Canadian federal NPM reforms reveal a distinctive 'Canadian' response. The federal government did not embrace the NPM practice of 'less government is better government.' Instead, it continued to take on an interventionist role and addressed issues "in terms of instrumental pragmatism" (Aucoin 2002:39).

\(^{13}\) New Public Management theory consists of a set of managerial and economic ideas and practices that were adapted from the private sector and implemented in the public sector.
Furthermore, the local implementation of NPM reforms in OECD member states varied and "followed a contextually-defined logic" (Salskov-Iversen et al 2000:184).

Larner (2005) reminds us that neoliberalism is not necessarily a coherent project but has been combined with other rationalities in order to achieve various policy objectives. She notes that academics often conflate these rationalities:

Neoliberalism, on the one hand, involves a focus on market individualism. Neo-conservatism, on the other hand, involves moral authoritarianism. More recently what we might call ‘Third Wayism’ has tried to articulate neoliberalism to social democratic aspirations. We need to be more sensitive to the fact that there are different configurations of neoliberalism. Close inspection of particular neoliberal political projects is more likely to reveal a complex and hybrid imaginary, rather than the straightforward implementation of a unified and coherent political philosophy (11).

In examining OECD education policy, one observes that it has been branded with different political rationalities. Under Keynesianism, educational policy formulations were constituted in terms of public investment in education and state planning. These policies resulted in the expansion of secondary and post-secondary educational systems among OECD member states. Optimal management of educational resources and equality of education were high on the policy agenda.14

Under neoliberalism, OECD education policy focused on implementing accountability and performance measures, improving educational quality and monitoring of educational systems. As state-run institutions, educational

---

14 Keynesianism emphasized government intervention in the economy and the regulation of international capital flows.
systems came under criticism for their poor performance. This was particularly the case for American schools which were blamed for the poor performance of American students in international student assessments in the 1980s. Government officials argued for more reliable measures of student achievement to assess future American competitiveness in the world economy. I will describe how officials from the U.S. Department of Education approached the OECD and insisted on the creation of a system of international educational indicators. The American model\(^\text{15}\) for educational accountability that was adopted by the OECD provided a set of educational indicators that could be used by American policymakers and politicians to compare themselves to other industrialized countries. At the same time, Department of Education officials began to fund an international student assessment for science and mathematics that was to be implemented by the International Association for the Evaluation of Educational Achievement (IEA). Paralleling these developments, there was a growing interest in the linkages between adult literacy skills and human capital formation which culminated in the implementation of the first International Adult Literacy Survey (IALS) in 1994.

In the 1980s and early 1990s, neoliberal international policy advice was guided by the logic of the market (Porter and Craig 2004; Deacon 2005). This logic has been reconfigured in the late 1990s to embody the Third Way social investment state (Lister 2004; Porter and Craig 2004). The shift can be characterized as a movement towards third way politics where individuals

\(^{15}\) I use the term American model in a similar way as Graefe (2006) does – to describe "policies and standards that the USA is championing in the international fora" (199).
cannot be “abandoned to sink or swim in an economic whirlpool” (Giddens 1998:100). Instead, a new mixed economy needs to be created where there is a “synergy between public and private sectors, utilizing the dynamism of markets but with the public interest in mind” (101). More specifically, third way advocates promote investment in education and lifelong learning because of their “capacity to improve both economic efficiency and social cohesiveness” (Latham 2001:29).

In the area of global social policy, Deacon notes that “the intellectual tide is turning against the neoliberal social policy prescriptions arguing everywhere at a national level for targeted benefits only for the poor” (2005: 25). A recent review of several international reports shows that poverty, social inequality and equity are back on the global agenda (Koivusalo 2006; Noël 2006). As this study will show, the OECD’s educational policy formulations have shifted as the neoliberal project has mutated. The OECD PISA is a reflection of this “complex and hybrid imaginary” (Larner 2005: 11) that embodies “a striking coexistence of technocratic economic management and invasive social policies” (Peck and Tickell 2002: 389). Under Third Wayism, the OECD’s approach to education policy includes maintaining accountability and performance systems, measuring the quality of educational systems, and monitoring the performance of schools. This approach, however, is also accompanied with concerns for inequities and the monitoring of social inclusion/exclusion of marginalized groups in educational systems.
### Table 1: Political Economic Context and Education Policy Developments

<table>
<thead>
<tr>
<th>Political Rationality</th>
<th>Time Period</th>
<th>Political Economic Context</th>
<th>Education Policy Developments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keynesianism</td>
<td>1940s – 1970s</td>
<td>Bretton Woods (1994): Regulated flows of capital and fixed exchange rates. Government intervention in the economy: maintain high levels of employment by stimulating aggregate demand and low levels of inflation by ensuring wages do not fall.</td>
<td>Public investment in education and state planning; expansion of secondary and post-secondary systems of education; creation of comparative student assessment techniques; institutionalization of the IEA.</td>
</tr>
<tr>
<td>Neoliberalism and Neo-conservatism</td>
<td>1970s – early 1990s</td>
<td>Breakdown of Bretton Woods: Deregulation of international flows of capital and flexible exchange rates. Minimal government intervention in the economy; market individualism; retrenchment of the welfare state; and, adoption of private sector managerial techniques in the public policy sphere.</td>
<td>Implementation of accountability and performance measures; improvement of educational quality; privatization of education; testing of adult literacy skills; creation of the OECD’s International Educational Indicators Project; implementation of the IEA’s international student assessment for testing students in science and mathematics.</td>
</tr>
<tr>
<td>Neoliberalism and Third Wayism</td>
<td>1990s – present</td>
<td>Maintain neoliberal economic order that privileges market individualism and globalisation of trade and capital; allow for social inclusion policies that develop social and human capital.</td>
<td>Maintain accountability and performance systems; continue to measure quality of educational systems; concern for inequities in educational system; creation of the OECD PISA to measure students’ literacy skills in a knowledge economy.</td>
</tr>
</tbody>
</table>

Before turning to the sociological theories that inform my theoretical framework, I want to first discuss the way in which international organizations fit...
into the global architecture of education and the role of American actors within this structure. These multilateral networks are crucial to distributing neoliberal ideas and knowledge that then become applied at the national and local levels. The Global Architecture of Education and American Dominance in International Relations

The global architecture of education is a concept I have borrowed from Phillip Jones who has used it to describe a “system of power relations that exerts a heavy, indeed determining, influence on how education is constructed around the world” (Jones 2006: 43). Jones examines how the World Bank has positioned itself as part of the global structure of education. By tracing the OECD’s role in the construction of the PISA, I also reveal how the OECD has positioned itself in an international statistical and data gathering and production infrastructure that exists within a global structure of education. In this section, I identify several international organizations involved in educational measurement and data gathering practices. I will then discuss the influential role that American officials have played in influencing the governance and policies of these international organizations.

The World Bank\textsuperscript{16} and the International Monetary Foundation (IMF) are multilateral economic institutions that propagate the “ideals of economic globalization” (Jones 2006: 49). They invest in infrastructural educational projects. The World Bank assesses and monitors these investments by using international student assessments to gauge progress. The OECD complements

\textsuperscript{16} The World Bank is a UN Specialized Agency.
the work of the World Bank and the IMF by spreading free market ideas and, as noted earlier, acts as a policy forum and policy instrument.\footnote{Another organizational actor that helps to put into practice ideas of free trade is the World Trade Organization (WTO). Educational services are covered under the WTO’s General Agreement of Trade in Services (GATS).}

The United Nations has several agencies that are heavily involved in education. These include the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations Children’s Fund\footnote{Originally known as the United Nations International Children’s Emergency Fund.} (UNICEF) and the United Nations Development Program (UNDP) (Jones 2006). For the purposes of this study, I am concerned with UNESCO’s work in education. UNESCO has worked closely with the OECD in creating a classificatory system for educational programs and in developing educational indicators.

Jones (1990), who has also studied UNESCO’s policies on literacy, notes that unlike the World Bank and the UNDP, UNESCO experiences a “paucity of resources” to implement its educational ideals (58). UNESCO was greatly weakened when the Reagan Administration withdrew from UNESCO in 1984. The reasons for the Administration’s withdrawal included the “heavily politicized” nature of UNESCO activities, the “compulsively statist” approach to UNESCO policy, and poor budgetary and management practices (Jacobson 1984: 582-583).\footnote{According to the New York Times, “UNESCO ... had, in Washington’s view, become the worst of the [United Nation’s] independent agencies. Encouraged by the Soviet Union and radical third-world countries hostile to the United States, the Paris-based body pressed for such antidemocratic programs as tighter government controls over journalists, as embodied in the plan for a “new world information order” (Bernstein 1984:6).} The George W. Bush Administration rejoined UNESCO in 2003 after a series of reforms were instituted.
There are several UNESCO organizations that are important to the analysis of the OECD PISA. As we shall see in Chapter 4, it was under the auspices of the UNESCO Institute for Education (UIE) that the first international student assessment was conceived.\(^{20}\) The project spurred the creation of the International Association for the Evaluation of Educational Achievement (IEA). The IEA, as will be discussed, is a rival organization to the OECD PISA. A second relevant UNESCO organization is its Institute for Statistics. This institute was created in 1999 when UNESCO’s statistical programs were reformed. The Institute for Statistics conducts educational surveys and produces its own educational statistical indicators. A third UNESCO agency is the International Institute for Educational Planning (IIEP) which was created in 1963. IIEP activities include what it describes as “monitoring educational quality.” The IIEP has helped transfer knowledge of standardized measures of educational achievement and was involved with the creation of a regional student assessment organization known as the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ).

An important regional organization in the educational global infrastructure is the European Union (EU). The European Commission, the executive body of the EU, has observer status at the OECD. The European Commission’s statistical agency, EUROSTAT, produces educational statistics and gathers educational data collaboratively with the OECD and UNESCO. Since the 1990s, the European Commission and the European Council, represented by the

\(^{20}\) UNESCO announced in 2006 that the UIE will become the Institute for Lifelong Learning which “is more closely aligned to the Institute’s objectives and activities” (UNESCO 2006).
Heads of State, have taken an active interest in national educational policy directions and emphasized the need for a competitive and skilled workforce in a knowledge-based society (Jones and Duceux 2006; Mitchell 2006).

American governments have attempted to dominate the international policy sphere since World War II and have influenced policy directions within international organizations. At times, they have used multilateralism as an instrument for perpetuating American interests and for maintaining a certain international political economic order conducive to American prosperity and capitalist growth (Karns and Mingst 1990; Woods 2002; Foot, MarcFarlane and Mastanduno 2003). American power and influence is reflected in the "Americanization" of transnational governance. My dissertation reaffirms what Djelic and Sahlin-Andersson have claimed – that there exists "a unique and often powerful role and place of American actors and blueprints in regulatory processes, both at the origins and at critical key moments." They note, "[t]his power and influence is particularly linked historically to the post Second World War period and is associated in part with the threading of an international organizational net - key nodes being the World Bank, the IMF, the OECD, the United Nations and its satellites, the GATT and the WTO" (2006: 397).

The global infrastructure for educational measurement initially reflected an American model for governing education – one that was influenced by the ideas of the Republican Party under Ronald Reagan and George H. W. Bush. Since its inception, the American educational model has been reconfigured. This reconfiguration occurred under the Clinton Administration as it adopted a
Third Way approach to education policy by “investing more and demanding more” from American schools. The approach combined “standards and accountability” with “investment in proven strategies” (“President Clinton touts education reform agenda” 2000). It has also occurred through a series of negotiated compromises with OECD member states to reflect a more Europeanized understanding of educational quality and human capital formation – one that emphasizes the social investment role of education. This hybridized approach to education policy may not address the objectives of the George W. Bush Administration since American educational reforms have moved towards even more stringent accountability standards reflected in the No Child Left Behind Act which was signed into law in 2002 (McGuinn 2006).

The OECD PISA as a Capacity-Communication-Power Bloc

I apply Foucault’s capacity-communication-power bloc to analyze how the OECD PISA exercises power. This framework enables me to structure my analysis in terms of the OECD PISA’s ability to build its technical capacity, to develop relations of communication and to build global relations of power. As Foucault explains:

> In a given society there is no general type of equilibrium between finalized activities, systems of communication, and power relations. Rather there are diverse forms, diverse places, diverse circumstances or occasions in which these interrelationships establish themselves according to a specific model. But there are also “blocs” in which the adjustment of abilities, the resources of communication, and power relations constitute regulated and concerted systems.

My analysis traces how this adjustment has taken place by examining three types of relations – relations that build technical capacity, relations of
communication and relations of power. These relationships “always overlap one another, support one another reciprocally, and use each other mutually as means to an end” (1983: 218). In order to be able to undertake such an analysis, I have separated these relationships, but it is important to note that they are inter-related.

(a) Capacity Building: Communities of Practice and Social Worlds

The OECD PISA’s technical capacity is built through the growth of communities of experts who occupy various social worlds and who are involved in educational measurement at the national and international levels. Members of these communities develop new practices and tools as they acquire knowledge in their fields. I use the term communities of practice (Lave and Wenger 1991) to describe members of learning communities who are involved in international student assessment work and who are taking part in the “worlding operation” of international standardized educational measurement practices (Gane and Haraway 2006: 139).21 Community members are also members of social worlds, “groups with shared commitments to certain activities sharing resources of many kinds to achieve their goals” (Clarke 1990). Social worlds intersect when their work activities come to be problematized along similar lines. They also can segment as new problem areas are construed. In studying the emergence of the PISA, I trace the emergence of new social worlds, the

---

21 Researchers in the policy field use the term epistemic communities to describe international “networks of knowledge-based experts” and to explain international policy coordination (Haas 1992: 2). I prefer the term communities of practice to epistemic communities because it can incorporate expert and non-expert participants and it can be applied across different scales and across different forms of social and bureaucratic institutions.
intersection of social worlds, and their segmentation (Clarke 1990; Strauss 1978).

Scientific communities build technical capacity as they define the problem and seek to resolve it. Devices are used by community members to interest other members in the problematization. These new members want to be involved in resolving the problem. As intéressement takes place, new members are enrolled. Enrolment ensures that allies are locked into place. In other words, it becomes difficult for new members to leave the community (Callon 1986).

I use the concepts of “doability” and “tinkering” to analyze the way in which community members resolve educational measurement problems and build their technical capacities (Fujimura 1987). Doability is attained through the presence of certain technologies and tools – for example, statistical techniques, computers and software. Tinkering involves changing the organization of work and practices. Tinkering means that educational researchers and measurement experts “make problems doable through the seemingly mundane processes of organizing and reorganizing their work” (Fujimura 1987: 258). I study how the science for measuring student achievement was rendered doable.

(b) Relations of Communication: Infrastructure, Heterogeneous Engineering, Boundary Objects, Standardized Packages, and Immutable Mobiles

I have described my approach to analyzing the capacity building relations of international student assessments. In order for experts to continue building on
these capacities, they require an infrastructure for communicating with one
another as they render their work more doable through the creation of new tools
and work activities and as they stabilize facts.

Members of communities of practice are linked together through relations
of communication. They create an infrastructure for communicating with one
another across institutional structures, geographic boundaries and levels of
government. Community members deploy various tools to facilitate
communication. These communication activities and acquired aptitudes
"produce effects of power" that are connected to a global architecture of
education (Foucault 1983: 218).

Infrastructure is a concept used to describe the PISA as a negotiated
order that creates tools for its users, matches work practices with available
resources, and is comprised of negotiated compromises (Bowker and Star
1999). Infrastructures facilitate the exchange of knowledge, ideas, tools, and
practices and encompass relations of work (Star 2002). These relations are not
readily visible unless an individual is involved in work activities and is a member
of a community of practice. I investigate the work practices of the PISA
members and its users and the compromises and adjustments that have been
made to enable the infrastructure to exist. Convergence is achieved as
members construct tools that fit with their work activities. By convergence, I
mean that the infrastructure was stabilized through the construction of tools and
artefacts that fitted a community's work practices (Bowker and Star 1999).
To understand how relations of communication are formed, we need to be able to think in terms of human and non-human entities or actors. Infrastructures are created through a process of heterogeneous engineering that shapes and assimilates human and non-human entities into a network of associations (Law 1987:113). Actors build worlds by associating with other entities (Callon 1991).

I use the concepts of methods of standardization and boundary objects to analyze how organizational actors from different social worlds are linked together and come together to resolve problems of coherence and cooperation. In the case of the construction and implementation of the PISA, methods of standardization are implemented through stringent quality standards and regulations to managing information across social worlds.

Boundary objects help coordinate work across social worlds by being customizable. They can be tailored to address local needs. They are "simultaneously concrete and abstract, specific and general, conventionalized and customized … and often internally heterogeneous" (Star and Griesemer 1989: 408). They are "weakly structured in common use" and "strongly structured in individual-site use" (393). An example of a boundary object used in constructing the PISA is a generalized definition of literacy that facilitates work across social worlds (weakly structured in common use) and provides simultaneously for a customization of its meaning at the local level (strongly structured in individual–site use). Another example of a boundary object is a classification system. As we shall see in Chapter 5, classification systems allow
for the coordination of work across communities of practice because they are customizable and can be tailored to address local needs (Bowker and Star 1999). They can be molded because they are weakly structured in common use. At the same time, they are strongly structured in individual-site use because they are used for local classification work.

I use the concept of standardized package to examine an object that “handles both collective work across divergent social worlds and fact stabilization” (Fujimura 1992). Boundary objects are combined with standardized methods so as to “narrow the range of possible actions and practices” without “entirely defin[ing] them” (169). These standardized packages serve as “interfaces” between intersecting social worlds (178). They provide “variability among worlds and specificity within work sites” (187). The PISA domain frameworks act as standardized packages. They do boundary object work as well as standardizing work. They contain a flexible conceptual framework and domain definition on the one hand and a more standardized methodology for developing test items on the other hand. The OECD PISA domain frameworks have the characteristics of standardized packages containing both boundary objects and standardized methods.

Relations of communication are facilitated through the deployment of a variety of inscriptions or immutable mobiles. These “representational tools” (Roth and Barton 2004: 32) such as maps, drawings, graphs, numbers, money,

---

22 Domain frameworks are the conceptual maps for defining the knowledge being assessed, the test construction criteria and the interpretation of results for each of the domains. Each domain - reading literacy, mathematics literacy and science literacy – has its own domain framework.
and data are "mobile, flat, reproducible, still and of varying scales" and "can be reshuffled and recombined" (Latour 1986: 21). What is important to note about immutable mobiles is their ability to grow as they gather a large number of inscriptions from one location and carry these inscriptions to another, more centralized location.

Immutable mobiles travel across the global architecture of education. They are crucial to the exercise of power as they help forge relations of power among human and non-human entities. Immutable mobiles are deployed to carry information back and forth from the periphery (schools) to the centres of calculation located in various multilateral organizations and agencies such as the OECD and the UNESCO Institute for Statistics.

(c) Power Relations
To analyze the relationships of power within a power bloc formation, I examine how the PISA governs the conduct of individuals and schools. I study how the PISA instrument legitimizes what counts as knowledge through the codification and measurement of an object called 'literacy.' I examine the dominant institutional structures that reproduce the knowledge and the politics inside these institutions.

I also examine global relations of power. In order to 'see' better, international organizations monitor the quality of educational systems and gather educational indicators. Knowledge is accumulated at a centre of calculation through action at a distance (Latour 1987: 219). That is, knowledge is accumulated of distant places, students, and educational systems and brought
"home" to be reshuffled and recombined to produce new knowledge, which can serve as the basis of further local interventions.

**Studying the OECD PISA**

The dissertation examines the development of the PISA's technical capacity, relations of communication and power relations. Chapter 3 begins by providing an overview of major developments that helped create a science of educational measurement and contributed to building the technical capacity for implementing large scale student assessments. It focuses on how an object called 'intelligence' came to be quantified and how intelligence testing was rendered doable. Chapter 4 examines how technical capacity was built to conduct standardized testing internationally. It explains how members from different social worlds problematized the need for international indicators of student achievement in the late 1950s. I describe how the International Association for the Evaluation of Educational Achievement (IEA) was formed and trace the conflicts and debates that accompanied its various studies of student achievement. Chapter 5 traces OECD educational activities and focuses on how it built its capacity for gathering educational statistics.

Chapter 6 examines how American government officials have used the OECD, as they have used other international economic organizations, to "legitimize and universalize" their national interests (Foot *et al* 2003: 4). Their level of activism has varied, depending on American domestic politics and

---

23 Wade (2002) has examined American influence on the World Bank; Foot, MacFarlane and Mastanduno (2003) have edited a book which contains articles analyzing American international multilateral cooperation and influence.
priorities and the international political environment (Foot et al. 2003; Karns and Mingst 1990). American influence within multilateral organizations has been contested and was never fully hegemonic. The research led me to see more clearly that American officials with the Department of Education played a vital role in problematizing the absence of international educational indicators.

Chapter 7 describes how the OECD PISA was formed. The chapter discusses the political maneuverings and negotiated compromises that underpinned the construction of the PISA. I discuss the formation of a PISA community of practice and examine its tools and its work practices. I analyze the way work was coordinated through the deployment of boundary objects, standardized packages and immutable mobiles.

Chapter 8 examines the relations of power in which the PISA is implicated. I discuss how the PISA instrument quantifies and codifies an object called ‘literacy.’ I examine how the knowledge produced by the PISA instrument becomes official knowledge and thus attempts to govern the conduct of individuals. I also examine the institutional power relations that exist within the OECD and the International Association for the Evaluation of Educational Achievement (IEA).

Chapter 9 illustrates how the PISA is used by various end-users and how it is an integral part of a global architecture of education. It returns to the theme of the fragility of an international science for educational measurement and focuses on the vulnerability of the PISA. This vulnerability is due to the historically contested discipline of international student testing and the politicized

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
environment that exists within the OECD. Comparing the performance of students cross-culturally is hard work that is imbued with politics. It requires constant negotiation of areas of debate and conflict.

The PISA can only exist if the artefacts and knowledge it produces continue to be used by communities of users ranging from educational researchers, statisticians, psychometricians to policymakers and politicians. More significantly, the legitimacy and relevance of the PISA depends on the support it gets from ministries of education and government agencies, especially from American officials who continue to see the merit for using the PISA data to inform domestic educational objectives. It is the connection of the PISA to a global educational network that further legitimizes the knowledge it produces and facilitates the distribution of this knowledge.

In the next section, I discuss the data gathering methods that I used during the course of my research.

Data Gathering Methods
The study draws from a range of primary sources, including OECD documents and conference proceedings, OECD International Educational Indicators Project meeting minutes and newsletters, OECD PISA Board of Participating Countries/Governing Board meeting minutes, and OECD and IEA documents circulated during meetings and newsletters. IEA, OECD PISA and other OECD publications were also consulted. Meeting minutes were available on-line. Some documents were provided to me by individuals whom I interviewed.
contacted the OECD Education Directorate in order to obtain unpublished PISA-related documents. The Directorate provided me with a zipped file with most of the documents I requested, except for those that were less than three years old.

In reconstructing the history of international student assessments, journal articles written in the 1950s, 1960s and 1970s were consulted extensively. It was more difficult to reconstruct the history of the OECD PISA. Other than the OECD discourse that existed on the PISA, there was little written on its actual creation. In many cases, the meeting minutes did not reflect the politics involved in the OECD PISA's creation. I had to rely largely for this information on twenty-two interview respondents whose accounts represented insider histories. Where possible, I tried to corroborate the accuracy of the information that was provided to me with other interviewees and with the meeting minutes.

Twenty-two interviews were conducted with various senior officials, middle-managers and experts involved in the construction of international educational indicators, assessments, and statistics. The twenty-third interview respondent was an academic from a German university. Ten of the interviews were with experts who were involved with the PISA Functional Expert Groups (Reading, Mathematics and Science), the PISA Technical Advisory Group and the Australian Council for Educational Research. Their names and the organizations they were affiliated with were listed in OECD PISA publications. Two senior OECD officials who were involved with the development of OECD educational indicators and with the Education Directorate were also interviewed. As well, three delegates from the PISA Governing Board (formerly known as the
PISA Board of Participating Countries) participated in the study. Two senior officials from the U.S. Department of Education were interviewed in order to obtain a historical understanding of American involvement in international educational assessments. Three senior officials who were involved with the IEA and with the UNESCO Institute for Statistics participated in the study. Two co-directors from the IEA's International Study Center at Boston College were invited to participate in the study but they declined to be interviewed.

The interviews supplemented the documentation and archival material I collected. In certain cases, I had to rely on interviews as my primary research source since there was no other written record that documented the history I was tracing.

Most of the interviews were conducted over the telephone, since many of these individuals were located around the globe. A few interviews were face-to-face interviews. During the telephone interviews, I took notes and sent the discussion notes to each interview respondent for his or her review. The interview participant would then email me back the notes with his/her revisions. Certain participants in the study provided me with answers to follow up questions. Face-to-face interviews were recorded and then transcribed.

All interviews took place in compliance with Carleton University's Ethics Committee regulations. Interview respondents were sent an information letter describing the research work and a consent form which they signed and faxed.

24 Although I would have preferred to conduct personal interviews, I did not have the financial resources to travel internationally.
25 In retrospect, I should have purchased equipment that recorded telephone interviews in order to transcribe the interview discussion accurately.
back to me. The information that the interview respondents provided remained anonymous and opinions and responses were not attributed to any single individual. Interviewees were concerned with confidentiality and anonymity particularly since they were associated with a very high profile project under the auspices of a powerful organization, the OECD. I have done my best to ensure complete confidentiality despite the fact that the interviews were drawn from a relatively small sample.\textsuperscript{26} I cite the interviews by a coded number and date.

Conclusion

My analysis of the OECD PISA uses a multi-disciplinary approach to trace how the PISA was constructed, how it exercises power and how it is connected to global networks of power. The political economic context situates the PISA within the broader political rationality of neoliberalism. The international relations lens provides me with the tools to explain American influence in the PISA's construction and more generally, in the governance of international organizations. Sociological theories, particularly those of Michel Foucault, help frame my analysis of the OECD PISA as a power bloc formation. I integrate concepts from the sociology of science and technology in order to analyze how the OECD PISA exercises power as a bloc of capacity-communication-power.

The OECD PISA is analyzed as an integral component in the global architecture of education, connecting communities of practice across levels of governance and across international organizational structures. The PISA

\textsuperscript{26} Appendix 1 contains a list of interviews and dates; Appendix 2 contains a sample Consent Letter that the interview respondents signed; and, Appendix 3 contains a list of questions that the interview respondents were sent so as to help guide the discussion.
becomes one node in a global network that constructs official knowledge and material practices. As we shall see, the analysis points to the fragility of the PISA. Despite the political and scientific work that has informed the PISA's construction, it remains vulnerable. I claim that the controversies that surrounded the field of comparative international education continue to haunt the PISA. The power bloc formation of technical capacity, relations of communication and power relations as a regulated system of inter-relationships is constantly being weakened.

In the next chapter, I present an overview of major developments that helped create a science of educational measurement and how technical capacity was built to quantify an object called 'intelligence.'
Introduction

The chapter provides an overview of major developments that helped create a science of educational measurement and contributed to the technical capacity of such large scale assessments as the OECD Programme for International Student Assessment (PISA). The developments discussed in this chapter include the construction of 'intelligence' and its measurement; the creation of a scale for diagnosing intelligence; the measurement of an individual's IQ; the creation of the group test; the development of new testing technologies; and the evolution of testing theory.

The historical overview demonstrates that the quantification of intelligence occurred as eugenicists, educational researchers, statisticians, and psychometricians rendered this science doable. The construction of a quantifiable object called intelligence was not an easy process to undertake and required significant tinkering by members of these communities of practice. The chapter also discusses the controversies and debates that arose as the science of educational measurement was being constructed. The controversies have never been fully resolved and, as we shall see in later chapters, they were compounded once they were taken from the national to the international levels, contributing to the fragility of programs such as the OECD PISA.
The Construction of ‘Intelligence’ and its Measurement

It was Sir Francis Galton who, in his book *Hereditary Genius* (1869), attributed intelligence to heredity (White 2000). Galton named the eugenics field and defined it as “the science of improving [human] stock, ... which takes cognizance of all influences that in however remote a degree give to the more suitable races or strains of blood a better chance of prevailing over the less suitable than they otherwise would have had” (Galton, quoted in Fancher 1985: 28). He set out to support his hypothesis by demonstrating a “statistical tendency for eminence” that was hereditary. Galton wanted to find out how good and bad traits were “preserved by heredity” (Desrosières 1998: 128). Preservation of traits was pivotal for Galton’s Eugenics Society that was to be founded by two parents who scored high on a series of examinations.

Galton is viewed as the “father of psychometrics” – “the measurement of quantitative behavioral traits” (Jensen 2002: 148). Galton developed the first idea of the intelligence test when he set up his Anthropometric Laboratory at the International Health Exhibition at London’s South Kensington Museum in 1884. Galton collected data from nine thousand paying visitors to the Laboratory who underwent a series of sensory and motor skills tests that determined their natural abilities (Fancher 1985; Jensen 2002). Galton wanted to prove that a link existed between intelligence and sensory discrimination (Acton 2001).

To analyze his research data, Galton experimented with the technique of regression and with the mathematical idea of the correlation coefficient. It was Karl Pearson, Galton’s student, who, twenty years later, turned Galton’s
eugenics concept of regression and correlation into a mathematical form (Desrosières 1998). Pearson’s student, Charles Spearman, further elaborated on the application of Galton’s notion of general mental ability by discovering a general factor of intelligence that he called “g” (Jensen 2002; Sacks 2000).

Spearman discovered “g” through a series of mathematical and statistical techniques that analyzed data compiled from school children’s abilities in certain school subjects, rank in school and teachers’ ratings. Spearman published his findings in 1904 in *The American Journal of Psychology* in an article entitled, ‘General Intelligence,’ in which he presented his conception of intelligence as a “unidimensional and inherited general ability” (Fancher 1985:98). To this date, educational psychologists and psychometricians have debated whether such a thing as “g” exists or whether it is a statistical construct (Buckalt 2002).

Spearman’s research utilized a new statistical technique called factor analysis. His contribution shifted the debates over the measurement of intelligence away from whether it was possible to measure such psychological traits to the methodologies that were to be used to quantify these attributes:

By mathematically reducing a complex set of correlations to a small number of dimensions, or factors, factor analysis provided a statistical means of determining what it was that intelligence tests were measuring. In so doing, factor analysis enabled the basic theoretical question underlying mental testing to be reduced to a technical difficulty (Hornstein 1988: 12).

27 In statistics, correlation describes a relationship between two variables. A correlation coefficient is used to express this relationship in mathematical terms. If the relationship is perfectly positive, the correlation coefficient will be 1.00. If the relationship is perfectly negative, it will be -1.00. If there is a relationship, then the correlation coefficient will be zero. If two variables have a positive relationship, the coefficients will have a value between zero and 1.00 (Borg & Gall 1989: 574).
For psychometricians, the construct "g" is central to the quantitative research of intelligence. It is useful to see how psychometricians conceptualize "g" in their application of factor analysis since it forms the "central dogma of psychometrics" (Buckalt 2002: 102):

**Figure 1: A hierarchical factor model of 'g'**

In Figure 1, the factor model depicts Spearman's general factor of intelligence, "g," as composed of a hierarchical structure of abilities. At Level 1 of the hierarchy are single test items that measure a particular ability or competency. There may be a series of items that measure that one ability. These test items are not correlated. At Level 2, test developers create a series of tests consisting of Level 1 items that are more correlated than at Level 1. At Level 3, the first order factors are derived from Level 2 correlations. At Level 4, a smaller number of second order factors are derived from Level 3 correlations. Finally, at Level 5, a single third order factor is derived, "g," representing the "highest order common factor in a large battery of diverse cognitive tests" (Jensen 2002: 152).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
There were problems with Spearman's research because his methods seemed to conflate performance and intelligence.

Behind the seemingly elegant statistics, the most glaring problem with Spearman's experiments on abilities was this: His independent measures of intelligence were in fact proxies for the very same abilities he was assessing. In other words, his intelligence measures—tied as they were to performance in the school subjects and teachers' opinions—were by definition dependent on changes in his chosen abilities. They amounted to the very same thing, and so of course, they would seem to correlate highly with each other (Sacks 2000: 22).

The techniques used in Spearman's work—reliability testing, correlation coefficient and factor analysis—came to be widely used, however, in all types of measures of ability including standardized tests, examination results or teachers' ratings (Cattell and Butcher 1968:14). Spearman's methods gave "quantitative precision" to Galton's work (Sacks 2000: 19). What was still missing from the science of educational measurement was a practical application of an intelligence test that provided a measure of "psychological magnitude" (Hornstein 1988: 10).

A Scale for Diagnosing Intelligence

Alfred Binet, a French psychologist, expanded on the practice of testing the senses as a measure of mental ability and set out to measure higher intellectual abilities. In 1903, he published his findings in *L'Étude Experimentale de l'Intelligence*. In it he discussed various one-on-one testing techniques that stimulated a person's intellectual and reasoning capacities (Binet 1903). His research led him to conclude that there were mature and immature intellects
with the mature intellects having a higher capacity for abstraction (Fancher 1985).

Binet found that statistics could not reflect the complexities that arise in the case study experimental method and did not advocate their use for measuring intelligence.

Mere numbers cannot bring out ... the intimate essence of the experiment. This conviction comes naturally when one watches a subject at work... What things can happen! What reflections, what remarks, what feelings or, on the other hand, what blind automatism, what absence of ideas! ... The experimenter judges what may be going on in [the subject's] mind, and certainly feels difficulty in expressing all the oscillations of a thought in a simple, brutal number, which can have only a deceptive precision. How, in fact, could it sum up what would need several pages of description!

We feel it necessary to insist that the suggestibility of a person cannot be expressed entirely in a number, even if the latter should correspond exactly to his degree of suggestibility. It is necessary to complete this number by a description of all the little facts that complete the physiognomy of the experiment (cited in Fancher 1985:63).

Binet's approach was defined as individual psychology and required one-on-one interaction between the subject and the psychologist. Even though his approach for measuring intelligence differed from methods deployed by statisticians, the underlying premise for his research remained a stable notion of an object called intelligence that could be classified (Garrison 2001).

With the enactment of universal education laws by the French Ministry of Public Instruction, all children were required to have several years of public instruction. These laws included providing special classes and programs for children with mental disabilities. The challenge arose as to how to sort legitimately the 'normal' and the 'retarded' children. In 1904, a commission was
appointed “to investigate the state of the mentally subnormal in France” and Binet was invited to become a member. As commissioner, he learned that educators were in dire need of a reliable and useful diagnostic system. Although three broad categories were being informally used to categorize institutionally the ‘mentally retarded,’ consisting of ‘idiots’, ‘imbeciles’ and ‘débiles’, an official classificatory system for differentiating among these categories was not in place (Fancher 1985, Garrison 2001).

Alfred Binet and Theodore Simon, who was doing his doctoral research under Binet’s supervision, set out to develop the criteria for diagnosing mental retardation. They developed what came to be known as the Binet-Simon Scale, which consisted of 30 tasks or subtests involving memory use, imagination and comprehension. Children were tested on an individual basis for 40 minutes. A child’s ‘intellectual level’ or ‘mental age’ was determined based on the percentage of children of a given age who arrived at the right answer.

The ‘right’ answers reflected bourgeois cultural attitudes that were usually learned in the home. When the scales were administered, both Binet and Simon observed that children from wealthier families performed better than poorer children (Hornstein 1988, Sacks 2000). Binet was not bothered by such outcomes since they validated the eugenics theory that class differences are hereditary:

That this difference exists one might suspect because our personal investigations, as well as those of many others, have demonstrated that children of the poorer class are shorter, weigh less, have smaller heads and slighter muscular force, than a child of the upper class; they less often reach the high school; they are more often behind in their studies (cited in Spring 1972: 6).
Today, test practitioners have developed the technical capacity to 'control' for this type of cultural bias and have created complex measures for testing validity. Nonetheless, the problem of culturally biased tests remains since tests are political projects reflecting dominant values and attitudes. As we shall later discuss, the OECD PISA has been criticized for being culturally biased.

The significance of the Binet-Simon Scale is that it "provided a set of age norms for each task, which meant that a given child could be compared with others, either of the same age or of a different age" (Hornstein 1988: 10). It was deployed as a technique for differentiating among groups of children. In reality, the Binet-Simon Scale legitimated the categorizations that were already institutionally being practiced (Garrison 2001).

The Creation of the Intelligence Quotient

The Binet-Simon Scale was translated into English by Henry Goddard, an avid eugenicist and professor of psychology. Its use spread among American educational practitioners and psychologists. When it came under criticism for being an inadequate measure of 'feeble-mindedness,' Lewis M. Terman, who was Professor of Educational Psychology at Stanford University, defended Binet's work and emphasized the value of Binet's contribution: "It is no reflection upon other workers to say that until very recently the Binet system of tests was the only ready means of making even an approximate estimate of the mental

---

28 Feeble-mindedness was a term used to refer to individuals with mental or learning disabilities.
level of a subject. Many other good tests were floating around, but because they had not been standardized they were of very limited value" (1916: 531).

The introduction of a standardized method for assessing the psychological magnitude of a person's intelligence led researchers and psychologists to develop a simpler representation of a person's intellectual ability. One such representation was the Intelligence Quotient or IQ.

The IQ was first conceived by the German psychologist, William Stern. Stern came up with a way of measuring intelligence results as a single number he called the Mental Quotient. It was a number calculated from the results from two sets of studies. The first study involved using the Binet-Simon scale to test children and then to re-test them a year later. Stern found that there was a gap between the children's mental and chronological ages (the test score being the indicator for mental age). His second study involved understanding these gaps. He found that both chronological age and "severity of their diagnosis" needed to be taken into account. Stern concluded that "measuring backwardness" needs to be done by calculating the "relative difference" or "the ratio between mental and chronological age" (Stern, quoted in Fancher 1985:103).

Lewis M. Terman Americanized the Binet-Simon Scale by expanding and revising it. His scale measured all levels of intelligence rather than simply identifying whether an individual is 'feeble-minded' or not (Lagemann 2000). His revisions included incorporating William Stern's calculation of the Mental Quotient, but he used the term Intelligence Quotient, which he abbreviated to IQ. His IQ classification system was composed of 7 categories – with below 70
being the lowest IQ – that of feeble-mindedness and an IQ of above 140 being near genius.

- IQ above 140: “Near genius or genius.”
- 120-140: “Very superior intelligence.”
- 110-120: “Superior intelligence.”
- 90-110: “Normal or average intelligence.”
- 80-90: “Dullness, rarely classifiable as feeble-mindedness.”
- 70-80: “Border-line deficiency, sometimes classifiable as dullness, often as feeble-mindedness.”

Terman conducted his own research in order to estimate the number of American children who fell into each category. His results indicated that there was a correlation between IQ scores and social class. He disregarded this test bias in his research and insisted that the reason certain individuals scored higher was because “their heredity was better” (cited in Sacks 2000: 28).

The invention of the IQ facilitated the reporting of results on intelligence tests, but the administration of such tests remained cumbersome and lengthy. The assessment took forty-five minutes to an hour and required a one-on-one assessment conducted by a professional psychologist. It was not a very efficient method for assessing the intelligence of large numbers of individuals. In the next section, we will discuss how the administration of intelligence tests was drastically changed.
From Testing Individuals to Testing Groups: The Creation of the Group Test

Robert Yerkes, the president of the American Psychological Association, was instrumental in transforming the Binet-Simon and Stanford-Binet scales from an individual test that required a professional psychologist to administer the assessment to a group test that could be administrated to a large number of people. This change in practice was needed in order to make intelligence testing a useful tool for military use. The army required an efficient selection process to classify effectively the huge influx of recruits as it prepared to enter World War I (Carson 1993; Spring 1972).

In 1916, the National Academy of Sciences created a Psychology Committee of the National Research Council "to mobilize America's scientific resources for defense" (Kevles 1968: 567). Chaired by Yerkes, the committee members included some of the most well-known psychologists at the time; Walter Bingham, Henry Goddard, Thomas Haines, Lewis Terman, Frederic Wells, and Guy Whipple.29 The committee members developed a test instrument for army recruits known as Army a (Carson 1993; Kevles 1968).

As a test instrument, Army a was designed differently from civilian intelligence tests. The instrument eliminated the need for a professional psychologist who administered the test and assessed the subject's behavior.

29 Carson (1998) notes that several psychologists declined Yerkes' invitation because they disagreed with his approach. They found it to be impractical, too "mechanized", and were worried that his quantitative approach would not provide sufficient information. To address these criticisms, Yerkes modified the assessment and "added a new feature that marked the next step in the evolution of Army a: the correlation of the psychological measurements with the industrial and military history of the individuals examined" (see p. 278-288).
The psychologist was replaced with a quantitatively designed instrument that produced a single number reflecting the subject's level of intelligence. The psychologist's role became interpreting the data that were produced by the test instrument. This new instrument succeeded in "remaking mental testing into a new sort of endeavor, one in which professional judgment was subordinated to objective determination and statistical manipulation" (Carson 1983: 287).

Army a was tested on 4,000 military personnel and the results analyzed. The Committee developed several recommendations based on its analysis of the Army a data that were then presented to the Surgeon General of the Army, William Gorgas. The Committee's recommendations included that "a first line army be organized; that men of inferior ability be reserved for later use or classified appropriately in the second or third line" (cited in Kevles 1968: 568). In July 1917, Gorgas assigned Major Yerkes to be the chief of a newly created psychological section in his office. Yerkes also became a member of the Committee on the Classification of Personnel that was formed with the army's Office of the Adjutant General.

Members of the Committee on the Classification of Personnel produced two sets of tests — the Army Alpha for literate subjects and the Beta for illiterates (Chapman 1988; Fancher 1985; Lagemann 2000). The Army Alpha tests consisted of eight parts and included such tests as "number sequences, word analogies, arithmetic problems, synonym-antonym puzzles, and commonsense questions." The Army Beta tests were "pictorial examinations, like the comparison of forms and the completion of partial drawings" (Kevles 1968: 572).
An infrastructure was built to administer, grade, and analyze the test results and subsequently to distribute the army recruits according to their assessment results among the various companies. It included establishing a school of military psychology for training test examiners, setting up army camps where the tests were administered and then graded, and creating a process for distributing army recruits so that each company received its share of "superior men, average men, and inferior men." This infrastructure, which was very effective in administering large-scale assessments, tested 200,000 men per month. In fact, the "Committee on Classification of Personnel managed to rate over 3 million men, some 1,750,000 of whom took the Alpha and Beta tests" (Kevles 1968: 573).

The army tests were not limited to identifying individuals at the two ends of the intelligence spectrum – that is the 'feeble-minded' and the 'geniuses.' Even though civilian psychologists were beginning to examine the "center of the intelligence spectrum," they knew very little about individuals in the middle. The army tests changed this practice. Military psychologists analyzed the data from recruits who were at the center of the intelligence scale and slotted them into the appropriate placement within the military organization (Carson 1983: 294).

Even though the Alpha and Beta tests were deemed to be useful in assisting with the army selection process, they were culturally and racially biased instruments (Sacks 2000, Fancher 1985). For instance, army recruits who were more 'Americanized' outperformed recent immigrants on the tests. There were questions on the tests that only someone who had experience living
in an American environment could answer correctly. For example, on the Beta test, pictures were presented to the test taker that reflected American cultural contexts such as bowling alleys, mailing letters, tennis players playing a game of tennis. Carl Brigham,\(^{30}\) who was a member of Yerkes' committee, concluded in his 1923 book, *A Study of American Intelligence*, the results "show clearly that the foreign born are intellectually inferior to the native born" (cited in Sacks 2000: 29). In reality, the army tests correlated with army officers' rankings, thus reinforcing institutional selection practices (Spring 1972: 8).

Once the war was over, the practice of group testing spread quickly in American educational institutions (Chapman 1988; Kevles 1968). Among the contributing factors for the popularization of the group test was the widespread availability of the tests. After the war, the army sold unused Alpha and Beta tests at lower prices than other commercially available tests. These were quickly taken up by educational practitioners and psychologists (Spring 1972: 4). The perception of intelligence testing changed after the war. Instead of testing being viewed as solely for identifying the feebleminded, it came to be viewed as applicable to everyone (Lagemann 2000: 92). The army tests rendered the practice of intelligence testing into a "respectable" practice (Kevles 1968: 580).

The 'respectable' practice of intelligence testing has led to a form of scientific racism. The argument that intelligence is hereditary has been used to legitimize the implementation of discriminatory and racist policies such as the

\(^{30}\) Brigham would later create the first college entrance exam, the Scholastic Aptitude Test (SAT).
sterilization of the feeble-minded, the disabled, and racial minorities (Dennis 1995; Dikötter 1998).

In short, thus far, intelligence research has advanced the politics of oppression, further stigmatizing and marginalizing minority individuals to justify policies and strategies that focus on elitism and exclusion instead of on meritocracy, egalitarianism, and equal opportunity (Onwuegbuzie and Daley 2001: 218).

Intelligence testing reinforced cultural and racial differences by culturally and racially grouping results. Psychometricians have highlighted these differences, pointing to the poor performance of African-Americans as compared to White Americans on IQ tests (Buckalt 2002).

As the use of intelligence tests spread in schools, they legitimated racist school practices that sorted and categorized children. Testing students' mental aptitudes became essential to the functioning of educational bureaucracies. Intelligence testing “reconciled equality of opportunity with inequality” so that the data confirmed the “naturalness of social class and social caste” (Reed 1987:77). In reality, “intelligence testing was a new “rational” scientific way of establishing meritocracy – a caste system based on “IQ”” (Kramer and Johnson 1997: 41). Students were classified according to their mental abilities on intelligence tests without taking into consideration their economic and social circumstances and the context of their daily lives.

In the next section, I discuss the impact of new technologies on the proliferation of student testing in American educational institutions.
New Testing Technologies

New technologies contributed to the proliferation of student testing in American educational institutions. The invention of multiple choice test items and true or false test items helped “transform the testees’ answers from highly variable, often idiosyncratic ... into easily marked choices among fixed alternatives” (Samelson 1987:116). The construction of such test items required a more specific definition of which skills were to be assessed. For example, the “essence of reading ability ... was measured by the rate of categorization of words” where students are tested on their ability to “find and then underline all words that fitted a particular concept” (117-118). These test items were deemed to be more objective since they standardized achievement and intelligence tests. Multiple choice and true or false test items were easy to score and did not require knowledge of the subject matter being tested. After appearing in 1914 on state level reading achievement tests, multiple choice testing became quite common by 1921 in intelligence testing and in army testing (Samelson 1987).

On the technological front, the invention of a machine that could correct multiple choice tests helped further spread the use of testing in schools and the spread of large-scale testing. Since mass testing of American students was quite resource and labour intensive and likely to be error-prone, a machine was needed instantly to correct these standardized tests. In 1931, Reynold B. Johnson, a high school science teacher in Ironwood, Michigan, invented an electrical test scorer he named the Markograph. Johnson improved on his Markograph so that it could electrically sense graphite inscribed answer sheets.
IBM bought the rights of the Markograph from Johnson and marketed its own machine (Lemann 1995, 1999). In 1937, the IBM Type 805 International Test Scoring Machine was introduced to the market (IBM 2007).

The invention of new technologies assisted in rendering the science of educational measurement more doable. These technologies helped further build the technical capacity of implementing large-scale standardized testing projects and facilitated the process of making them doable internationally.

Another important contributor to building the technical capacity has been the growth of theory in the educational measurement discipline discussed in the next section.

The Evolution of Testing Theory

Theoretical frameworks help to facilitate relations of communication among community members and contribute to establishing power relations as exclusive knowledge is produced and practiced among groups of experts. There are two test theories that predominate in the educational measurement field. The first is Classical Test Theory (CTT) and the second is Item Response Theory (IRT). CTT assumes that a single underlying trait or dimension can have one single true score, T. All test items measure this one trait. So, the test score, X, is calculated based on one’s true score, T, plus some error of measurement, e. The equation for CTT is depicted as: X=T+e.

IRT is similar to CTT in that it also measures a single underlying trait or dimension. The difference between the two is that “the basic conceptual idea
behind IRT is the relationship between the score (usually right/wrong) on an item and the underlying dimension, whereas CTT is based on the relationship between total test score and the underlying dimension" or trait. With the development of Item Response Theory (IRT), tests can be customized and adapted to each test taker. The error of measurement, e, varies for each test taker.

In Item Response Theory (IRT), the relationship between the score of an item and the trait it measures is portrayed by an item characteristic curve, or ICC. The probability of getting an item right changes with the underlying trait (Rindskopf 2004:13026).

**Figure 2: An example of an Item Characteristic Curve**

![Item Characteristic Curve Diagram]

Adapted from Wright and Stone (1979: 14) and Interview Respondent #7, 9 June, 2006.

There are different parameters that can be used in Item Response Theory (IRT). These include the item's difficulty, the discrimination of an item or the degree to which the item discriminates between persons, and the effects of guessing (for
example on a multiple choice question). One-parameter models or the Rasch model use only difficulty with guessing fixed at zero and discrimination set to one for all items. Two-parameter item response models use discrimination and difficulty with guessing fixed at zero. Three-parameter models are the most complex and take into consideration difficulty, discrimination and guessing. Item behavior will differ depending on the number of parameters used. In one-parameter models, item responses will move up and down the line representing the point of discrimination. In two-parameter and three-parameter models, item behavior tends to move along a range on the curve (Rindskopf 2004:13026; Interview Respondent #7, 9 June, 2006).

Even though for the layperson these may not be significant differences, for psychometricians and statisticians different parameter applications reflect different schools of thought. According to one interview respondent, the school of thought that is adopted is almost "like a religion" (Interview Respondent #15, 29 August 2006). In the educational measurement field, evaluation agencies apply different models. For example, Educational Testing Service and Statistics Canada frequently apply two-parameter models. The one-parameter model or the Rasch model is used more frequently by European and Australian statistical and evaluation agencies.

Georg Rasch, a Danish statistician, developed the one-parameter model in the 1950s. His model was derived from work he had done on reading progress of dyslexic children for the Danish Ministry of Social Affairs in 1951. He later adapted his model to test the intelligence of army recruits for the Danish

Benjamin Wright, the Director of the Measurement, Evaluation, Statistics, and Assessment Laboratory (MESA lab) at the University of Chicago, was very much inspired by Rash's work and helped to popularize Rasch's model among American academics (Olsen 2003). For Wright, Rasch's contribution "marks the point at which psychometrics moved from being purely descriptive to become a science of objective measurement" (Wright 1980: ix). Wright explains that objective measurement is "measurement that transcends the measuring instrument." It "function(s) independently of the objects measured" (ibid). It also is "a response model for calibrating their functioning, which can separate instrument and object effects." For Wright, this can be achieved by the use of parameters. As Wright explains in his Foreward to Rasch's seminal work, Probabilistic models for some intelligence and attainment tests: "It was the Danish mathematician, George Rasch, who first understood the possibilities for objectivity that reside in the logistic response model with one item parameter and one person parameter, and it was Rasch who first applied this model to the analysis of mental test data" (ibid).

Geoffrey Masters, who was to become the Australian Council for Educational Research's (ACER) CEO, studied under Benjamin Wright. He, in turn, continued with the application of Rasch's work initially begun at the
Measurement, Evaluation, Statistics, and Assessment Laboratory (MESA lab) at ACER (Wright 1979). One of Masters’ students, Ray Adams, has become a leading expert in Rasch measurement methodology as well. Adams was to become the Project Director of the PISA Consortium. In later chapters, I will discuss how ACER came to lead the Consortium\(^{31}\) that won the bid for the first PISA cycle.

In discussing the development of testing theory, I have provided a glimpse of the complexity of statistical knowledge that communities of practice in the areas of psychometrics and statistics have attained. The technical capacity of practitioners has grown as have the sub-specializations within the field.

**Conclusion**

This chapter discussed the major developments that contributed to building a science of educational measurement. It brought out the main debates and controversies surrounding the field of educational measurement and the theoretical foundations underpinning testing theories. I argued that these controversies, which would later undermine the stability of the practice of assessing students internationally, shed light on the fragility of such a practice.

The chapter examined one set of relationships of the power bloc formation – that of technical activities and capacity. I examined how members of communities of practice built the technical capacity for testing intelligence and

\(^{31}\) Other members of the International Consortium include the Netherlands National Institute for Educational Measurement (CITO), Westat from the United States, the Education Testing Service in the United States (ETS) and the Japanese National Institute for Educational Research (NIER).
showed how they rendered the science doable as they experimented with new
techniques and implemented new technologies. I discussed the way in which
they tinkered with the science of educational measurement, inventing new
practices and processes as they tried to quantify intelligence. Even though
these community members tried to make their science factual, they faced
constantly the problem of quantifying a psychological concept that was not a
physical property.

I pointed to the controversies that arose as psychometricians and
psychologists quantified intelligence. The existence of a construct such as “g”
was questioned. The racial and cultural biases of intelligence tests were difficult
to eliminate since they reflected the test creators’ prejudices and values. I
identified how among psychometricians, there were disagreements on the
application of testing theories. These issues would later arise in international
student assessments. In the next chapter, I trace how technical capacity was
built for an empirical science of comparative education and the founding of the
International Association for the Evaluation of Educational Achievement.
Introduction

In the previous chapter, I highlighted major developments in the educational measurement field and pointed to the antagonisms that are inherent in the testing discipline. In Chapter 4, I investigate how the capacity was built to conduct standardized testing internationally. To accomplish this objective, I trace how the International Association for the Evaluation of Educational Achievement (IEA) was founded and how it developed new practices and tools as it pursued its activities. We see how members of the IEA rendered their science more doable as they pursued their activities and developed their tools. I examine the controversies that arose as this community constructed a science of international student assessment.

Through this narrative, we begin to see the formation of a power bloc as capacity is built, relations of communication are established and relations of power among international organizations are instituted. The science of comparative educational measurement, and more specifically international student achievement testing, was used to inform policy makers on the effectiveness of schools and on possible reforms they could pursue. The application of new statistical techniques such as Item Response Theory would greatly assist in the reporting of results as ministries of education needed more complex measures to address student performance in an era of accountability. As we shall see, neoliberal educational reforms influenced the direction of
international student assessments as the IEA studies tried to report indicators that closely addressed these accountability requirements.

The Formation of the IEA

An international student assessment community of practice was formed in the 1950s as a result of a series of meetings that brought together educational researchers and sociologists from various educational research centres. This community of practice came to be represented by the International Association for the Evaluation of Educational Achievement (IEA). Today, the IEA is a non-governmental organization that relies on a vast and decentralized infrastructure of research institutes, universities and ministries of education to fulfill its work in international large scale assessments. The IEA Secretariat is based in Amsterdam, the Netherlands and its Data Processing and Research Center is located in Hamburg, Germany. The IEA’s contribution to the field and knowledge of international assessment and evaluations is considered to be immense and unprecedented. It has trained some of the world’s leading statisticians and psychometricians in international testing (Bottani and Vrignaud 2005).

Several developments contributed to the problematization of international student achievement. By problematization, I mean how comparing the performance of students internationally came to be defined as a problem and structured by those who founded the IEA. These developments can be situated in the 1950s as American funding in the field of comparative education rapidly
expanded. New comparative education centers were established,\textsuperscript{32} new journals for disseminating comparative education research were founded,\textsuperscript{33} and conferences in comparative education were being held worldwide, providing forums for knowledge exchange. Europeans were interested in this information exchange as they sought to learn new methods for measuring the effects of the expansion of secondary education and in learning new evaluation techniques from their American counterparts (Husén 1979: 373-374).

With the Russians launching Sputnik in 1957, American policymakers began to question the merits of their own educational system. As one editorial noted, the Sputnik shock “unleashed a veritable storm of comparisons of American education with foreign educational systems” (Bereday 1958: 1).\textsuperscript{34} Americans became concerned with comparing their schools with European and Soviet ones. Even though educational reforms had been initiated in the early 1950s, Sputnik expedited the implementation of these reforms to ensure American schools produced more scientists and engineers (Bybee 1997). In this context, American comparative educationists felt that there was a need for a ‘scientifically’ based approach to studying education comparatively rather than a cultural or narrative approach “largely concerned with the exchange and collation of descriptive material” (Postlethwaite 1966: 356).

\textsuperscript{32} The Comparative Education Society was established in 1956 and the Centre for Comparative Research at the University of Chicago was established in 1957.
\textsuperscript{33} Comparative Education Review was founded in 1957. The International Review of Education was relaunched in 1955 as a publication of the UNESCO Institute for Education. Prior to World War II it was known as the International Education Review.
\textsuperscript{34} Books that received media attention were Admiral Rickover’s (1959) \textit{Education and Freedom} and Paul Woodring’s (1958), \textit{A Fourth of a Nation}. Both advocated the streaming of American students.
Educational researchers began to meet in the 1950s to discuss issues related to student evaluation and problems facing educational systems (Husén and Postlethwaite 1996; Bottani and Vrignaud 2005). Most of these individuals were academics who specialized in sociology or educational psychology and who were affiliated with university research centers. Members were “drawn together” by Dr. William Douglas Wall, the Director of the National Foundation for Educational Research in England and Wales (NFER) (Purves 1987: 10). Wall had previously worked for the United Nations Educational, Scientific and Cultural Organization (UNESCO) as Head of Education and Child Development and had developed an important network of contacts (Griffiths 2003). Held on the premises of the UNESCO Institute for Education, the meetings were a forum for exchanging information cross-nationally and cross-culturally.

It was during a 1957 meeting devoted to educational evaluation issues that several members decided to meet in 1958 at the UNESCO Institute for Education to “consider the possibility of undertaking a study of measured outcomes and their determinants within and between systems of education.” Instead of relying on graduation rates as a measure of educational productivity, these educational researchers proposed to measure children’s learning which might “yield a very different ‘productivity’ measure” (Husén and Postlethwaite 1996:129). They hoped that this form of international comparative educational research across educational systems would reveal “important relationships that

---

35 The UNESCO Institute for Education (UIE) was established in 1951/52 in Hamburg, Germany (then the Federal Republic of Germany) and was initially tasked with post-war reform of the German educational system. The UIE’s mandate gradually shifted to include “international comparative education” and “East-West cooperation in educational research” (UNESCO 2005). In July 2006, the UIE changed its name to the UNESCO Institute for Lifelong Learning.
would escape detection within a single educational system" (IEA 2004). More specifically, they wanted empirical rather than qualitative evidence on the quality of educational systems in an era when Americans and Europeans were still recovering from the Sputnik shock (Foster 1991).

Wall, the Director of NFER, chaired the 1958 meeting that brought together prominent members from the international educational research community (Husén and Postlethwaite 1996; Bottani and Vrignaud 2005). In the next section, I trace the academic biographies of some of the individuals who were the founding members of an international student assessment community of practice. This helps to illuminate the commonality of interests for an empirically based approach to comparative education that they shared and their intellectual orientations. In this section, I highlight the development of human capital theory.

**Founding Members of the IEA**

C. Arnold Anderson was a key individual at the 1958 meeting. Trained as a sociologist, Anderson was appointed the first Director of the University of Chicago's Centre for Comparative Research in 1958. The Ford Foundation, a philanthropic organization interested in the "empirical approach" of the social sciences and the "practical use" of knowledge for policymakers (Pramar 2002:

---

36 Anderson in a 1973 interview states: "the International Study of Achievement was in a sense invented by me, soon also by a couple of others, and is slowly widening in prominence" (Anderson and Bowman 1973: 145).
16), donated $350,000 to the University of Chicago to create the Centre (Bereday 1957).  

In a 1957 memo submitted to the University of Chicago's faculty, Anderson listed seven research projects that the Centre for Comparative Research would undertake. Research project VI on the list was "inter- and intra-national comparisons of effectiveness of education" comprising "outcomes of formal schooling", "studies of validity, reliability and predictability of examinations of various systems" and "comparisons of teaching methods, coverage, and performance in selected fields ... to diminish the endless unsupported arguments about the ‘quality’ of different school systems.” Research project VII on Anderson's list (and the most extensive) was the “evaluation of education as investment in human capital and its marginal yields” (Anderson and Bowman 1973: 156-157).

Anderson's interest in the economics of education stemmed from associating with two individuals at the State University of Iowa at Ames in the early 1940s. The first individual was Theodore Schultz who was then Chair of the Department of Economics. Schultz would later move to the University of Chicago and become the Chair of the Department of Economics from 1946 to 1961. Schultz's contributions to economics were in the areas of economics of agriculture and human capital theory (Johnson 1999). The second individual

---

37 Pramar (2002) analyzes the knowledge networks created by philanthropic organizations such as the Ford Foundation as a form of intellectual hegemony. Building American-based knowledge through the funding of researchers, scholarships, research institutes and think tanks was important to the creation of this intellectual hegemony which emphasized instrumental use of knowledge and empirical research.

38 The field of economics of education entails applying economic analysis to educational issues and policies.
who influenced Anderson was Mary Jean Bowman. Bowman, who married Anderson, was an economist with the Chicago school. She also worked closely with Anderson at the Center for Comparative Education and was responsible for furthering research into the economics of education (Foster 1991).

The Center for Comparative Education was exposed to the developments in the economics field at the University of Chicago's Department of Economics, also known as the 'Chicago school' (Anderson and Bowman 1973). The economists of the Chicago school shared a common understanding of economic policy that favored a free market economy and limited government intervention in the market (Miller 1962). Out of this understanding of economics, economists such as Theodore Schultz and Gary Becker at the Chicago school nurtured a new approach to labour economics. They were not interested in studying the demand side of the labour market and that the variables had an impact on the demand side such as the structure of the labour market, wage differentials, and the determinants of income distribution. Instead, they turned their focus to the labour market supply side and explained all differences in wages and employment in terms of investment in education and training. They assumed that labour markets operated efficiently so they were not to be interfered with. It was up to those entering the labour market to invest in their human capital and to compete for jobs. What was ignored (and continues to be ignored) in this theoretical framework was that low wages, unemployment, and underemployment were caused by economic structures (see Bluestone 1977).
The human capital theorists explained worker productivity and economic growth in terms of investment in people’s education and training:

Schooling, a computer training course, expenditures on medical care, and lectures on the virtues of punctuality and honesty are capital too in the sense that they improve health, raise earnings, or add to a person’s appreciation of literature over much of his or her lifetime. Consequently, it is fully in keeping with the capital concept as traditionally defined to say that expenditures on education, training, medical care, etc., are investments in capital. However, these produce human, not physical or financial, capital because you cannot separate a person from his or her knowledge, skills, health, or values the way it is possible to move financial and physical assets while the owner stays put (Becker 1992: 85).


In 1975, Samuel Bowles and Herbert Gintis published a Marxist critique of human capital theory. They pointed to how it excluded an analysis that dealt with relations of production and the unequal power relationships that existed between employees and employers (81-82). They noted that human capital theory restricted labour market policy analysis and intervention to technical matters such as schooling, training programs, lifelong learning, and human resource development. They concluded that

This framework provides an elegant apology for almost any pattern of oppression or inequality (under capitalism, state socialism, or whatever),
for it ultimately attributes social or personal ills either to the shortcomings of individuals or the unavoidable technical requisites of production. It provides, in short, a good ideology for the defense of the status quo. But it is a poor science for understanding either the workings of the capitalist economy or the way towards an economic order more conducive to human happiness (82).

Torsten Husén was another key participant at the 1958 UNESCO meeting who would later play a significant role in institutionalizing the international student assessment field and disseminating knowledge on international student evaluation. Husén was trained as a psychologist and worked initially as a military psychologist for the Swedish army. He became Sweden's first professor in applied educational research in 1956 and was involved in organizing research on Sweden's comprehensive schools. Husén was associated with Anderson and was invited to lecture on Swedish school reform at the Center for Comparative Education. Husén would later work as an educational consultant with international organizations such as the International Institute for Educational Planning (IIEP), UNESCO Institute for Education, and the OECD (Postlethwaite 1993).

Several other American researchers from Columbia University's Teachers College were invited to the 1958 at the UNESCO Institute for Education. These included Edward Thorndike's\textsuperscript{39} son, Robert Thorndike from the Institute of Psychological Research. Thorndike's expertise was in psychometrics. In 1954, he co-published the Cognitive Ability Test that aimed at producing a profile of intelligence. Other academics from Teachers College were Wellesley Foshay

\textsuperscript{39} Edward Thorndike founded the field of educational psychology.
and Harry Passow. Foshay was an expert in curriculum development and Passow\textsuperscript{40} in developing educational planning for the gifted. Both were affiliated with the Horace Mann-Lincoln Institute\textsuperscript{41} which was part of Teachers College, Columbia University (Foshay, Thorndike, Hotyat, Pidgeon and Walker 1962).

Another prominent American educational psychologist who was invited to the meeting was Benjamin Bloom. Bloom had studied under Ralph Tyler and was influenced by Tyler's work on educational evaluation. Tyler viewed evaluation as a tool for evaluating student progress towards learning objectives (Wolf 1999). Bloom operationalized these learning objectives by developing a taxonomy of educational objectives based on a hierarchy of six levels of cognitive operations. This taxonomy facilitated the work involved in assessing student test outcomes. Published in 1956, \textit{Taxonomy of educational objectives: Handbook 1, the cognitive domain} was used widely by educational evaluators (Eisner 2000).

The group's additional members included David Walker, Director of the Scottish Council for Educational Research; Fernand Hotyat, Director of the Belgian Centre des Travaux de l'Institut Supérieur de Pédagogie de Moralanwelz; Gaston Mialaret,\textsuperscript{42} Professor of Psychology and Pedagogy with the University of Caen in France and President of the International Association

\textsuperscript{40} At the time, Passow was the Director of the Talented Youth Project. Talented or gifted youth in American schools were identified based on their IQ scores (Passow 1957, 1960).

\textsuperscript{41} Established in 1943 to improve school programs through research and experimentation, the Horace Mann-Lincoln Institute of School Experimentation was part of Teachers College, Columbia University (Passow 1957).

\textsuperscript{42} Gaston Mialaret, an educational psychologist, was recognized as "le père des sciences de l'éducation françaises" (the father of the French educational sciences) (Mabilon-Bonfils and Saadoun 2003). In the 1950s, Mialarat established France's first psychological experimental lab.
for Experimental Education of the French-speaking countries; and Walter Schultze, Director of the Federal Republic of Germany’s Hochschule für internationale pädagogische Forschung (Higher Institute for International Educational Research). Douglas Pidgeon, who was at the time Senior Research Officer with the National Foundation for Educational Research, was another participant with technical expertise (Foshay et al 1962, Purves 1987, Husén and Postlethwaite 1996).

The founding members agreed to conduct a pilot study in order to find out whether inferences could be made from a large scale student assessment and whether such a study was feasible (Foshay et al 1962, Purves 1987). Bloom and Foshay wrote the proposal for the pilot study which they submitted to American officials with the Office of Education and which was subsequently submitted to the Governing Board of the UNESCO Institute of Education. The Governing Board agreed to undertake and fund the study (Purves 1987, Husén and Postlethwaite 1996).

The Pilot Study

The original group of educational researchers who met in 1958 met twice again in 1959. They were joined by representatives from other educational research centres who were interested in participating in the study. At these meetings, they discussed issues such as “sampling, test construction, questionnaire items, timetables, and costs” (Purves 1987: 11). The pilot study involved the

---

43 Founded in 1952, it is known today as the German Institute for International Educational Research or DIPF. The DIPF analyzes German PISA results and has a representative, Dr. Eckhard Klieme, who is actively involved with PISA at the international level.
coordination and collaboration of twelve educational research centers.\textsuperscript{44} The study was administered from 1959-1961 with 9,918 13-year old students tested in reading comprehension, mathematics, science, and geography.\textsuperscript{45} A non-verbal intelligence test was also administered that supplemented the background information that was gathered on each child through questionnaires (Foshay \textit{et al} 1962).\textsuperscript{46} Several of the test and background data components of this first pilot study were to become part of later international student assessments (Bottani and Vrignau 2005).

The results were collected in 1961 and analyzed by Thorndike at Teachers’ College in New York. Thorndike was the chief test editor and primarily responsible for analyzing the international data. Thorndike analyzed patterns of achievement and within country and between country variances of student achievement. Subsequent studies that the IEA was to conduct also included within and between country variances (Foshay \textit{et al} 1962).\textsuperscript{47}

In 1962, the results of this pilot study were published under the auspices of the UNESCO Institute of Education.\textsuperscript{48} The study confirmed that such an international comparative study was feasible and that valid inferences could be

\textsuperscript{44} Participating research centers from the following states: Belgium, England, Finland, France, the Federal Republic of Germany, Israel, Poland, Scotland, Sweden, Switzerland, the United States and Yugoslavia
\textsuperscript{45} Thirteen year olds represent the “school population near a terminal point”, that is at the end of their schooling (Foshay \textit{et al} 1962: 7).
\textsuperscript{46} Background data included the student’s date of birth, sex, number of siblings, place in birth order, home language, location of home, years in school, attended or not attended kindergarten, size of class, father’s and mother’s education, interest of parent (much, moderate, little, or no), father’s and mother’s occupations, score on non-verbal intelligence test.
\textsuperscript{47} I note this because PISA data also analyze within and between variances but the tone of one of the PISA interview respondents implied that the OECD approach was original and innovative, when in fact this form of variance analysis has been practiced since the 1950s by the IEA.
\textsuperscript{48} Foshay \textit{et al} (1962) \textit{The Educational Achievements of Thirteen-Year Olds in Twelve Countries}.
drawn from the data of an international student assessment (Foshay et al. 1962, Postlethwaire 1966, Husén and Postlethwaite 1996). At the same time, several problems were reported in the final report of the pilot study which are still encountered today. These included the reliability of sampling procedures, translation problems and the validity of cross-cultural comparability (Foshay et al. 1962).

The first pilot study set the infrastructural base for future studies of international student achievement. It established a certain set of work practices and created tools for administering an international student assessment. The international costs of the study were funded by the UNESCO Institute of Education and the national costs were paid for by each participating research centre. A project director and a project coordinator were assigned. The tests were developed and then translated. Trial tests were created for pre-testing. When the pre-testing was completed, the tests were edited by the chief test editor. The finalized tests were circulated for approval among individuals at the UNESCO Institute for Education. Upon approval, the tests were administered by the national research centers that were participating in the pilot study. Coordination work involved collecting the test results, sending them to Thorndike in New York for analysis and then distributing the analyzed data to all the participants. Punch cards were used to store the data that Thorndike processed at Teachers’ College, Columbia University (Foshay et al. 1962; Husén 1979).

The work practices established for the first pilot study have remained somewhat the same in subsequent international student assessments. The
changes that did occur involved the expansion of the infrastructure to accommodate a larger number of participating centers and a larger sample of students being tested. Infrastructural expansion required more intensive coordination and negotiation work, additional artefacts and more advanced technologies.

First International Mathematics Study

The international student assessment community of practice launched its first major study in 1960. Because of the complexity involved in collecting, analyzing and publishing the data, the community created an organizational framework for the study and named it the Project for the Evaluation of Educational Achievement (IEA). In 1960, the twelve national research centres that were to participate in the IEA’s First International Mathematics Study formed a Council. Among the original Council members was the Australian Council for Educational Research, which would break away from the IEA in the 1990s. The Council appointed five Council members to be on the Standing Committee responsible for the study’s supervision with Douglas Wall as the committee’s Chair.

Tensions soon arose between the Director of the UNESCO Institute of Education (UIE), Saul Robinsohn, and Douglas Wall. Robinsohn was not pleased with the creation of an institutional structure that operated from within the UIE - it threatened his organizational control and drained the UIE's financial resources. Wall wanted the IEA to have a certain degree of autonomy. Wall and Robinsohn disagreed, resulting in Wall’s resignation as committee Chair and the appointment of Husén as Chair in 1962. These tensions were reduced
when the IEA received $250,000 in financial assistance from the U.S. Office of Education\textsuperscript{49} (Purves 1987, Husén 1995).

The First International Mathematics Study\textsuperscript{50} (originally named the International Study of Achievement in Mathematics) deployed more elaborate and refined work practices and tools that helped construct the infrastructure for international student assessments. The IEA hired Neville Postlethwaite,\textsuperscript{51} who managed the Coordinating Center, as Project Coordinator for the study. This role ensured that project coordination took place from one central location. A Test Construction Committee was formed that developed the study’s test instruments. The Test Construction Committee received reports from each national centre describing the educational system’s mathematics curriculum and suggestions for test items. Three experts in mathematics education advised the Committee on the development of test questions. Several test booklets were created and manuals for test administration were written. The study also collected information on ‘environmental’ factors that were gathered from two student questionnaires, a teacher questionnaire, a school questionnaire, and a National Case Study Questionnaire. All of the study’s data were collected at one central location for statistical analysis, the University of Chicago’s Computation Centre (Husén 1967).

The First International Mathematics Study report pointed to problems of comparability. Some of these were related to rendering mathematics curricula

\textsuperscript{49} The U.S. Department of Education was created in 1980.
\textsuperscript{50} For a complete list of IEA studies, see Table 3.
\textsuperscript{51} Postlethwaite had gained experience in the areas of measurement and testing at NFER (Husén 1995).
from different educational systems comparable. For example, not all students were taught the same mathematical concepts at each grade level. The test developers tried to resolve this problem by introducing a wide range of mathematical concepts that would cover as much of the students' learning as possible. Another problem the study's creators encountered was defining the target populations. If age was used to define the population, then not all students in a particular grade level would be represented. On the other hand, if grade level defined the target population then not all 13 year olds would be represented. To resolve this conundrum, the study's creators decided to have both populations included: (a) students who are 13.0 – 13.11 years old at the date of testing and (b) students at the grade level where the majority were aged 13.0 – 13.11 is found. Another population that the study targeted was students in their last year of schooling. Since the last year of schooling varied across educational systems, the Study's creators did not define the population's age or its grade level (Husén 1967). These problems of comparability have not been eliminated and constant tinkering and negotiation has to take place so as to resolve them.

The IEA community of practice was able to undertake the First International Mathematics Study during the 1960s because certain technologies were available to its members. Community members were able to store the answer sheet data on punch cards using the IBM 1230 Optical Mark Scoring Reader. This machine optically read the answer sheets and transferred them to the punch cards. Using IBM 7094 computers, the punch cards were then
transferred onto magnetic tapes. The data were then manipulated (edited, sorted and organized into records) and statistically analyzed. The data were also stored as a data bank. To facilitate access to the data, the IEA published a Data Bank Manual for its users (Husén 1967; Wolf 1968). Immutable mobiles, in the form of data, facilitated the creation of an international student assessment infrastructure. By immutable mobiles, I mean two-dimensional representations that could travel across distance and be recombined and reshuffled to produce new artefacts. Such work activities and tools enabled the collection of a large amount of data from peripheral national research centres at a central location.

*The Six Subject Study*

From 1966 to 1973, the IEA undertook the Six Subject Study. The Study tested students in science, reading comprehension, literature education, English as a foreign language, French as a foreign language, and civic education. The student populations that were assessed included 10-year-old students, 14-year old-students, and students in the final grade of secondary school. A total of twenty research centers participated in the Six Subject Study and, for the first time, students from developing countries were also assessed. Table 2 provides information on the student populations that were assessed and the research centers participating in the study.

Comparability problems were further exacerbated in the Six Subject Study. The test instruments and questionnaires could not address the cultural variability that existed across systems of education and student backgrounds. It is interesting to note that even Husén cautioned in his 'Foreward' to the Six
Subject Study's science study report, "When reading the tables which present findings from these analyses, the reader should also be very careful in making comparisons between the highly industrialized and the less industrialized countries" (1973: 13). This statement suggests that Husén may not have felt

Table 2: The Six Subject Study\textsuperscript{52}

<table>
<thead>
<tr>
<th>Six Subject Study</th>
<th>Student Population</th>
<th>Research Centers</th>
</tr>
</thead>
<tbody>
<tr>
<td>First International Science Study</td>
<td>10-year-old students, 14-year old students, and students in the final grade secondary school</td>
<td>Australia, Belgium (Flemish), Belgium (French), Chile, England, Germany (FRG), Finland, France, Hungary, India, Iran, Italy, Japan, Netherlands, New Zealand, Scotland, Sweden, Thailand, United States.</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>10-year-old students, 14-year old students, and students in the final grade secondary school</td>
<td>Belgium (Flemish), Belgium (French), Chile, England, Finland, Hungary, India, Iran, Israel, Italy, Netherlands, New Zealand, Scotland, Sweden, United States.</td>
</tr>
<tr>
<td>Civic Education</td>
<td>10-year-old students, 14-year old students, and students in the final grade secondary school</td>
<td>Federal Republic of Germany (FRG), Finland, Ireland, Israel, Italy, Netherlands, New Zealand, Sweden, United States, England, and Iran participated in the initial phase of the project.</td>
</tr>
<tr>
<td>Literature Education</td>
<td>14-year-old students and students in the final grade of secondary school</td>
<td>Belgium (Flemish), Belgium (French), Chile, England, Finland, Iran, Italy, New Zealand, Sweden, United States.</td>
</tr>
<tr>
<td>English as a Foreign Language</td>
<td>14-year-old students and students in the final grade of secondary school</td>
<td>Belgium (French), Chile, Federal Republic of Germany (FRG), Finland, Hungary, Israel, Netherlands, Sweden, Thailand.</td>
</tr>
<tr>
<td>French as a Foreign Language</td>
<td>14-year-old students and students in the final grade of secondary school</td>
<td>Chile, England, Netherlands, New Zealand, Romania, Scotland, Sweden, United States.</td>
</tr>
</tbody>
</table>

\textsuperscript{52} Adapted from information available on the IEA website, 'Completed Studies,' available at http://www.iea.nl/completed_studies.html, last accessed July 28, 2007.
that the data were reliable enough to make valid comparisons between countries. Thorndike, in the Six Subject Study's reading comprehension report, also noted the following:

Comparison of level of performance in different countries represents a rather minor component of the analysis of results from the present study: The 15 countries studied differ in so many ways with respect to language, culture, level of economic development and pattern of education that it is an almost hopeless task to try to isolate critical factors accounting for differences in achievement (Thorndike, cited in Downing and Dalrymple-Alford 1974-75: 215).

The fragility of the science of international student assessment can be seen as members of the IEA struggled to address problems of comparability among educational systems.

The First International Mathematics Study and the Six Subject Study were criticized for their lack of validity and reliability. One scathing critique came from Hans Freudenthal who was a leading expert in mathematics education. He scrutinized several of the studies, including the first mathematics study and the literature, science and reading assessments that were conducted as part of the Six Subject Study. In a lengthy journal article published in July 1975 in the journal, *Educational Studies in Mathematics*, Freudenthal pointed out systematically the weaknesses of the IEA studies, such as the exclusion of mathematics subject experts in constructing the mathematics tests; the lack of valid test instruments and population samples; the ambiguous use of variables; the lack of rigorous hypothesis testing; and errors in translation of the test items. He concluded that the data produced by these studies were unreliable.
According to Freudenthal, Husén (the Chairman of the IEA) sent him a letter indicating that the IEA would not respond to his criticism of its surveys. The reasons being that Freudenthal's article "transgressed the border of *ad hominem* argumentation" and that his criticism was outdated since the study was "planned almost fifteen years ago ... therefore it comes somewhat post festum" (Freudenthal 1976: 529). Husén did include with his letter a copy of IEA's reply, 'A Commentary from IEA on Dr. Freudenthal's Article,' that was published in December 1976. The IEA Commentary, which was written by Gilbert Peaker (formerly Chief Inspector of Schools in Mathematics in England), was cordial but did not address the criticisms that Freudenthal discussed. The Commentary concluded with the following remark: "IEA is duly grateful to Dr. Freudenthal for drawing attention to mistakes hitherto uncorrected. This gratitude is slightly tempered by regret that the article does not show some awareness either of the limitations under which the undertaking was carried out or of what was actually done" (Peaker 1976: 526).

Freudenthal, in turn, published a rejoinder to the IEA Commentary. He was not very happy with the IEA's reply since, according to Freudenthal, the Commentary misconstrued his critique. He noted, "[r]eading what he cited from my paper, one could be led to believe that I disagree with IEA about some points of strategy, whereas I attempted to prove that the whole undertaking was mistaken in principle, and badly conducted" (1976: 530). Freudenthal's critique was never fully addressed by the IEA as it embarked on more studies.
Second International Mathematics Study and Second International Science Study

In response to interest voiced by several states, the IEA Council decided to undertake two major studies in the mid-1970s (Travers 1977). In 1976, it launched the Second International Mathematics Study and in 1978, the Second International Science Study. To address problems of curricular comparability and criticisms of its earlier studies, the IEA tried to establish a more inclusive process for framework and test development. In order to develop a standardized approach to defining populations cross-culturally, the IEA studies used UNESCO's International Standard Classification of Education (ISCED).53

The mathematics conceptual framework was developed during a five day meeting that brought together mathematicians, mathematics educators and educational researchers. The framework used a 3-tiered model of curricular knowledge: intended curriculum representing intended outcomes and official syllabi; implemented curriculum encompassing what the students are taught in the classroom; and attained curriculum representing students' learning and attitudes. This model depended on greater national involvement in determining the test and questionnaire content. National Mathematics Committees that included mathematics educators, mathematics teachers, school inspectors, university mathematicians and educational researchers were regularly to advise the International Mathematics Committee on their respective curricula. Such interaction was only achieved with a few national committees. Problems arose

53 I will discuss the emergence of ISCED in greater detail in the next chapter on OECD educational activities.
because only English was used as the language of communication. Information was transmitted by mail – a very slow mechanism for exchanging information (Travers and Westbury 1989; Kilpatrick 1990).

As the IEA members pursued these studies, they refined and improved their methodological and conceptual approaches to assessing students, to constructing theories, to developing test items and questionnaires, to implementing assessments, and to statistical data analysis. By building relations of communication, they also strengthened their technical capacity and knowledge. IEA members created venues for exchanging and imparting knowledge among community members. These included on the job training, technical seminars, and conferences. The technical capacities and statistical knowledge of centers associated with the IEA studies were enhanced as the IEA expanded the scope of its work. International Coordination Centres that housed the studies and the International Coordinators who were based at these centres gained significant experience running a large scale international assessment project. These developments helped strengthen the technical capacity of a science for internationally assessing students.

The IEA produced knowledge and tools that fitted the needs of its community of users, enabling them to analyze the test data for national and international purposes. The IEA was very productive in disseminating and communicating its results. Between 1962 and 1974, over fifty IEA related articles and publications were published by members of this community of practice (IEA Publications 1974). It incorporated new informational technology
and tools that rendered its work more doable. The IEA used computers for data collection, storage, processing and analysis. Early on, procedures were put in place for storing test results and questionnaires on punch cards. Computer programs were used for organizing the data and for statistical analysis. They continued to develop and adapt to new technologies.

The IEA was initially dependent on the UNESCO Institute for Education for resources. Even though the IEA’s organizational structure was strengthened in 1967 when it was incorporated, the IEA never did achieve financial independence. Initially referred to as the International Project for the Evaluation of Educational Achievement, the IEA It came to be known as the International Association for the Evaluation of Educational Achievement (IEA) and in 1969, the IEA moved its headquarters from Hamburg to Stockholm.

The IEA relied on several financial sources to maintain the production of its studies. The grants it received often had strings attached and conditions set by the granting institutions (Husén 1979). The IEA tried not to yield to the pressures and demands imposed on it by national policy makers and to maintain academic autonomy in its approach to the study of international student achievement.

In the next section, we shall see how during the neoliberal era of educational accountability, the IEA was forced to reorganize and address the needs of policymakers, particularly the requirements of the U.S. Department of Education.
The Influence of Neoliberalism

In the 1970s and the 1980s, the IEA underwent significant expansion. Under Neville Postlethwaite's chairmanship, the number of IEA participating states increased from twenty-five to forty (Husén 1995). The IEA embarked on two major studies in the mid-1970s: the Second International Mathematics Study and the Second International Science Study. In the 1980s, it launched the Study of Reading Literacy.

In 1985, the IEA began working on the Study of Reading Literacy that assessed 9 and 14 year-old students in thirty-two educational systems. The study was conducted in 1991 and used an instrument that measured a student's literacy and reading comprehension. Neville Postlethwaite, who was the project's director, ran the study out of the University of Hamburg. Andreas Schleicher, who was originally hired as Data Manager, became the study's International Coordinator. Schleicher would later become responsible for the OECD Programme for International Student Assessment (PISA). Postlethwaite was aware that policymakers increasingly wanted data for decision-making purposes. In fact, one of the study's reports was dedicated to "identify[ing] some of the indicators concerning schools and their operation that an educational planner might address in order to transform less effective schools into more effective schools" (Postlethwaite and Ross 1992: 3). The report's data analysis singled out several indicators that influenced student achievement, such as student activities at home, school context, school characteristics, school resources, and school initiatives (Postlethwaite and Ross 1992; Elley 1994).
The IEA’s Standing Committee and General Assembly were aware of the increased interest in educational accountability and in international educational indicators that was being voiced by several member states and by international organizations such as the World Bank. Yet, because the IEA did not have a strong organizational structure, it was not equipped to address the reporting requirements for international educational indicators in educational achievement. In 1988, the IEA decided to formalize and institutionalize its organizational structure. It set up its permanent headquarters, the IEA Secretariat, in the Netherlands. In 1989, Tjeerd Plomp was appointed Chair of the IEA. Plomp was faced with the challenge of transforming the IEA from an organization run informally by academics to a professionally-run organization (Interview Respondent #9, 29 June, 2006).

Even though the IEA did become a more professionally run organization, it remained a non-governmental institution that neither had the political strength nor the financial resources to maintain its leadership position in the international student assessment arena. International organizations such as UNESCO were willing to collaborate with the IEA, but others, such as the OECD, preferred an arrangement in which more control could be exerted on “the ownership and timing of the data” generated by the IEA studies (OECD/INES 1995a:4). The IEA was criticized as an inappropriate venue for assessing students internationally. Its data collection system was deemed to be too fragmented and its data collection practices were not adapted to those of national educational systems (Owen, Hodgkinson and Tuijnman 1995). Dissatisfaction with the IEA
led to the problematization of a new international student assessment, the Programme for International Student Assessment.

The Adoption of New Testing Practices – Adjusting to the Accountability Requirements

The IEA Study of Reading Literacy used Item Response Theory (IRT) for constructing test items and for analyzing test results. As discussed in the previous chapter, IRT, and its particular application developed by the Danish statistician, Georg Rasch, were increasingly being applied by practitioners of educational measurement. IRT was perceived to be more flexible and accurate because a test item or a group of test items could be modeled to measure a specific ability. In contrast, Classical Test Theory (CTT) was modeled so as to measure one single trait based on a total test score.

IRT facilitated the development of adaptive tests since items could be organized according to grade levels, curriculum objectives and student abilities. These abilities were ranked or scaled according to levels of difficulty. Items that contained easier tasks were ranked at a lower scale than items consisting of more difficult tasks. Test items could be organized and stored in an item bank where items could be easily retrieved and reused and additional items could be deposited. IRT required more complex statistical modelling and better computerized systems for churning out complicated mathematical algorithms and for storing large amounts of data (Embretson 1999; Molenaar 1995; Rudner 1998). Because IRT was able to develop more complex analysis of assessment
results, it was regarded as a better tool to use for measuring performance of educational systems.

The Reading Literacy Study used the Rasch model so that a single measurement scale was constructed that took into account a student’s ability and item difficulty. For example, less difficult items that required lower levels of ability were scaled from five to eight and included reading tasks such as short simple stories. More difficult items that required higher levels of ability were scaled from seventeen to nineteen and involved reading long, complex stories. Student test results on test items were then statistically analyzed using multivariate analyses\(^{54}\) (Elley, Schleicher and Wagemaker 1994).

In the 1980s and 1990s, a few national assessment centres began to adopt Item Response Theory as their primary method for testing students. In certain institutions, the adoption of the IRT Rasch model was very controversial. This was especially the case at the National Foundation for Educational Research for England and Wales (NFER). NFER had created an Assessment Performance Unit in 1974 in order to collect data on whether students' learning was improving. Influenced by Bruce Choppin,\(^{55}\) Assistant Head of Research and one of the earliest proponents of the Rasch model, NFER's Assessment Performance Unit had originally adopted the Rasch model for its student assessments in order to monitor educational progress over time. NFER decided to drop the Rasch model when it came under criticism for undertaking such an

---

\(^{54}\) Multivariate analysis is a branch of statistics concerned with simultaneously investigating two or more variable characteristics which are measured over a set of objects. Factor analysis, for example, is a technique used in multivariate analysis (Kachigan 1982).

\(^{55}\) Choppin was a student of Benjamin Wright.
approach. Harvey Goldstein, a prominent statistician and a critic of the Rasch model, advised against its use (Choppin 1981). When NFER dropped the Rasch model, Choppin resigned in 1981. NFER hired Goldstein as a consultant (Griffiths 2003) who would later become one of the most vocal critics of the OECD PISA methodology. The Institute for Objective Measurement, an organization dedicated to disseminating knowledge on the Rasch model, described these events at NFER as “educational progress collid[ing] with the juggernaut of entrenched interests,” represented by the “British educational establishment” (Linacre 1995).

Other assessment centres readily adopted the Rasch model. For example, the U.S. National Assessment for Educational Progress (NAEP) incorporated IRT methods in 1983 (Elliott and Phillips 2004). The Australian Centre for Educational Research (ACER) was contracted to test elementary students in New South Wales in 1989. The entire assessment applied Rasch analysis. More specifically, ACER’s leading psychometrician, Geoffrey Masters, was a strong advocate of the Rasch approach (Doig 1990). We begin to see fissures in the IEA social world as community members developed and adopted new measurement approaches. The IEA community of practice was segmented when ACER decided to submit a proposal for implementing an OECD-based student assessment.

---

56 The Institute for Objective Measurement publishes the Journal of Applied Measurement described as the authoritative Journal for Rasch measurement research, a newsletter for community members, and an annual periodical, Popular Measurement.
<table>
<thead>
<tr>
<th>IEA STUDY</th>
<th>Year(s) of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilot Study</td>
<td>1959-1962</td>
</tr>
<tr>
<td>Project Coordinator (Technical): Foshay</td>
<td></td>
</tr>
<tr>
<td>FIMS</td>
<td>1963–1967</td>
</tr>
<tr>
<td>First International Mathematics Study</td>
<td></td>
</tr>
<tr>
<td>Project Leader: Husen</td>
<td></td>
</tr>
<tr>
<td>Six Subject Study</td>
<td>1968 - 1973</td>
</tr>
<tr>
<td>1 - First International Science Study (FISS)</td>
<td></td>
</tr>
<tr>
<td>2 – The Study of Reading Comprehension</td>
<td></td>
</tr>
<tr>
<td>3 – The Study of Literature Education</td>
<td></td>
</tr>
<tr>
<td>4 – The Study of English as a Foreign Language</td>
<td></td>
</tr>
<tr>
<td>5 – The Study of French as a Foreign Language</td>
<td></td>
</tr>
<tr>
<td>6 – The Civic Education Study</td>
<td></td>
</tr>
<tr>
<td>Project Leader: Kenneth Travers</td>
<td></td>
</tr>
<tr>
<td>SIMS - Second International Mathematics Study</td>
<td>1977-1981</td>
</tr>
<tr>
<td>SISS - Second International Science Study</td>
<td>1982-1986</td>
</tr>
<tr>
<td>Written Composition Study</td>
<td>1983-1988</td>
</tr>
<tr>
<td>The Reading Literacy Study</td>
<td>1985–1994</td>
</tr>
<tr>
<td>COMPED - The Computers in Education Study</td>
<td>1987–1993</td>
</tr>
<tr>
<td>The Languages Education Study</td>
<td>1993–1996</td>
</tr>
<tr>
<td>Third International Mathematics and Science Study</td>
<td></td>
</tr>
<tr>
<td>Second Information Technology in Education Study Module 1</td>
<td>1997–1999</td>
</tr>
<tr>
<td>SITES-M1</td>
<td></td>
</tr>
<tr>
<td>Third International Mathematics and Science Study Repeat</td>
<td>1997–2001</td>
</tr>
<tr>
<td>TIMSS-R 1999</td>
<td></td>
</tr>
<tr>
<td>Civic Education Study - CIVED</td>
<td>1994-2002</td>
</tr>
<tr>
<td>Pre-Primary Project - PPP</td>
<td>1986-2003</td>
</tr>
<tr>
<td>Second Information Technology in Education Study Module 2</td>
<td>1999-2002</td>
</tr>
<tr>
<td>SITES-M2</td>
<td></td>
</tr>
<tr>
<td>Progress in International Reading Literacy Study 2001</td>
<td>2000-2003</td>
</tr>
<tr>
<td>PIRLS 2001</td>
<td></td>
</tr>
<tr>
<td>Third International Mathematics and Science Study Repeat Video Project</td>
<td>1998-2004</td>
</tr>
<tr>
<td>TIMSS-R Video</td>
<td></td>
</tr>
<tr>
<td>Trends in Mathematics and Science Study 2003</td>
<td>2001-2004</td>
</tr>
<tr>
<td>TIMSS 2003</td>
<td></td>
</tr>
<tr>
<td>Progress in International Reading Literacy Study 2006</td>
<td>2004-2008</td>
</tr>
<tr>
<td>PIRLS 2006</td>
<td></td>
</tr>
<tr>
<td>TIMSS 2007</td>
<td></td>
</tr>
<tr>
<td>Trends in Mathematics and Science Study Advanced 2008</td>
<td>2006-2010</td>
</tr>
<tr>
<td>TIMSS Advanced 2008</td>
<td></td>
</tr>
</tbody>
</table>

Conclusion

Chapter 4 described how a science for international educational measurement was rendered doable and showed how American educational researchers influenced the direction of research in international comparative education settings. The creation of the International Association for the Evaluation of Educational Achievement (IEA) embodied the ideal for an empirically based science of comparative education that was influenced largely by the American model of educational research. The analysis brought out the controversies and debates that arose in the field of comparative educational measurement. They were never eliminated and in fact continue to contribute to the system's vulnerability.

The chapter traced how power was exercised through the building of technical capacity, relations of communication and relations of power for a science of international educational assessment. I discussed the formation of a community of practice for international student assessment that came to be known as the International Association for the Evaluation of Student Achievement (IEA). I mapped its activities, infrastructural developments and global relations. The IEA addressed strategically the technical training and knowledge requirements of its members. It expanded its technical capacity as it rendered the science of international student assessment more doable.

In the next chapter, I discuss OECD educational activities and the development of its capacity for gathering educational statistics.
Chapter 5
Educational Activities at the OECD

Introduction
In the previous chapter, I traced the creation of the International Association for the Evaluation of Educational Achievement (IEA) and the institutionalization of a science in international educational assessment. My focus in Chapter 5 is to trace the development of OECD technical capacity for gathering educational statistics. I do so by examining the evolution of OECD educational policy during two time periods – from the 1950s to the 1970s and from the 1970s to the 1990s. During the first period, a Keynesian approach or welfare capitalist approach dominated. This approach favoured government intervention in the name of full employment and government expenditures in social programs. The second period reflects a shift towards market-oriented ideas known as neoliberalism. As we shall see, the OECD’s technical capacity for educational measurement was also accompanied by the development of its relations of communication as the OECD built an infrastructure and coordinated activities in this area. The OECD’s educational activities were woven into regional and global educational networks.

Post-World War II Era
During the post-World War II era, the Bretton Woods Agreement\(^{58}\) came to represent the stable functioning of the international political economy.

---

\(^{58}\) In 1944, delegates from 44 states met at the United Nations Monetary and Financial Conference in Bretton Woods, New Hampshire to discuss the postwar recovery and monetary issues. The outcome of this meeting was the Bretton Woods Agreement that created a postwar international monetary system of convertible currencies, fixed exchange rates and free trade.
International organizations served as mechanisms for ensuring that states collaborated and coordinated policies militarily, politically and economically. This relatively stable international political economic order lasted until the 1970s. In this section, I discuss the original mandate of the OECD, its organizational structure and its educational activities during this era.

As discussed in Chapter 4, the Sputnik shock created interest in empirical methods for researching education and spurred the creation of an international student assessment community of practice. It also mobilized the Eisenhower Administration into a new Cold War strategy. The strategy included additional funding for scientific research and the creation of the National Science Foundation in 1950. The majority of the funding went increasingly to “a select number of research universities, special research laboratories, industries or new institutions such as the Rand Corporation” (Vinovskis 1995: 34-35).

The Americans were determined to produce a highly qualified scientific and technical cadre capable of beating that of the Russians. In 1958, Congress passed the National Defense Education Act (NDEA) allocating $240 million in federal educational funding. According to the Department of Education (2006),

[t]o help ensure that highly trained individuals would be available to help America compete with the Soviet Union in scientific and technical fields, the NDEA included support for loans to college students, the improvement of science, mathematics, and foreign language instruction in elementary and secondary schools, graduate fellowships, foreign language and area studies, and vocational-technical training.

Multilateralism was one of the instruments the Eisenhower Administration used to enroll European allies into its Cold War strategy to build a highly
qualified scientific and technical cadre. In 1958, it gave the OECD’s precursor, the Organisation for European Economic Cooperation (OEEC) a $500,000 grant in order to establish the Office for Scientific and Technical Personnel (OSTP). The work practices and the tools that were developed by the OSTP helped create the technical capacity and the infrastructure for future OECD work in the area of education.

The OEEC’s original mandate was to assist with European economic recovery. By 1960, it had completed its mandate. Instead of dissolving the institution, the OEEC was transformed into the OECD and given a new mandate – to promote economic growth, employment and world trade. Its membership was expanded to include the Americans and the Canadians. Twenty states originally signed the Convention on the Organisation for Economic Co-operation and Development on 14 December 1960. Since then, ten additional states have become members of the OECD. Accession to the organization was defined (and continues to be defined) in terms of sharing “basic values” of “an

59 The OECD’s predecessor, the Organisation for European Economic Cooperation (OEEC), was created after World War II. It was assigned the mandate of distributing $12 billion in American aid provided by the Marshall Plan from 1948-1952 (see Aubrey 1967, Wolfe 1993, and Papadopoulos 1994 for a more detailed history of the OEEC).

60 The twenty states which ratified the convention in 1961 were: Republic of Austria, the Kingdom of Belgium, Canada, the Kingdom of Denmark, the French Republic, the Federal Republic of Germany, the Kingdom of Greece, the Republic of Iceland, Ireland, the Italian Republic, the Grand Duchy of Luxembourg, the Kingdom of the Netherlands, the Kingdom of Norway, the Portuguese Republic, Spain, the Kingdom of Sweden, the Swiss Confederation, the Turkish Republic, the United Kingdom of Great Britain and Northern Ireland, and the United States of America;

open market economy, democratic pluralism and respect for human rights. More recently, the OECD has pursued an enlargement strategy recognizing that its future may be tied to becoming a more diverse organization (see OECD 2004d).

The OECD is organized into a Council, the supreme body in the OECD, committees and directorates. The Council has permanent delegates and is chaired by the Secretary General. Delegates from member states are also represented on various specialized committees. The directorates of the OECD employ civil servants who service the various committees but who are expected to be loyal to the organization. The OECD Secretariat manages and administers OECD work. The OECD Secretariat is headed by the Secretary-General who is assisted by a deputy and assistant secretaries (Aubrey 1967).

The OECD fulfills three broad functions – that of policy actor, policy forum and policy instrument. As a policy actor, the OECD is an international organization that wields considerable influence in the realm of policy ideas. As a policy forum, the OECD is a place for delegates from member states to come together and deliberate on the problems they face. As a policy instrument, the OECD becomes a tool for policymakers to bolster reforms or changes in their respective states (Lingard 2000).

During the Keynesian era, OECD economists and educational researchers defined educational issues in terms of manpower planning and

---

62 See OECD webpage, Becoming a Member of the OECD: the Accession: Process http://www.oecd.org/document/11/0,2340,en_2649_201185_1958091_1_1_1_1,00.html. Last visited 06/09/05.
63 Currently, the OECD has opened accession discussions with Chile, Estonia, Israel, the Russian Federation, and Slovenia (OECD 2007).
investment in education. It was understood that states needed to invest in their manpower since human capital contributed to economic growth. In the educational policy sphere, the OECD focused its efforts on the collection and harmonization of educational statistics for educational planners, developing mathematical models for forecasting manpower needs of the economy, building the knowledge base for the field of economics of education and conducting regular reviews of national systems of education.

OECD educational activities of the 1960s and the 1970s represented a "front-end model of educational policy." Policies were geared towards expanding educational opportunities and enrolment in secondary and post-secondary schools (Rubenson 1999:11). They were governed by an understanding that investment in education improved the quality of human capital. OECD work practices included conducting reviews of each member state's scientific education and technical training, which came to be known as Educational Policy Reviews; developing methods for surveying the current 'stock' of scientific and technical personnel and forecasting future long-term needs; encouraging and assisting member governments in their educational planning requirements, particularly in developing a coherent science policy; and taking a leadership role in developing a new mathematics and science curriculum. The Office for Scientific and Technical Personnel (OSTP) achieved its objectives by coordinating conferences that brought together experts in their respective fields; publishing reports, surveys and reviews; and by providing
advice and technical assistance to member states (Eide 1990; Papadopoulos 1994).

The OECD’s geopolitical expansion was accompanied by a greater focus on the linkages between investment in education and economic growth. During the 1950s and 1960s, both European and American economists identified education as an important factor that contributed to economic prosperity. In 1960, when Theodore Schultz introduced his theory of human capital to the American Economic Association, he elaborated on the need for economics to treat “human resources” as a “form of capital, as a produced means of production, as the product of investment”. He noted that “[m]any of the paradoxes and puzzles about our dynamic, growing economy can be resolved once human investment is taken into account” (1961:3).

The Office for Scientific and Technical Personnel (OSTP) continued with its activities as the OECD’s Committee for Scientific and Technical Personnel (CSTP). The CSTP expanded its work in the area of economics of education by establishing a Study Group in the Economics of Education that regularly met between 1962 and 1965. Conferences were organized and reports published so as to “provide evidence about the positive contribution of education to economic growth” (OECD 1980a). Although the Study Group was initially academically-oriented, it became more policy-focused in areas such as educational planning and financing education for economic growth. Once OECD member states had

---

64 The CSTP became the Education Committee in 1970 (Papadopoulos 1994:23).
built their educational planning capacity, the Study Group was no longer needed and its work ended in 1965 (Papadopoulous 1994: 42).

In the mid-1960s, the Committee for Scientific and Technical Personnel began to develop a "network of mathematical model builders." This network created a set of measurement tools to forecast future student enrolment and to analyze quantitatively the efficient use of inputs for educational planning purposes. Mathematical macro-models were created, but these were too complex and could not be used locally by educational planners, school administrators and educational policymakers. The tools were created by a group of mathematicians who were not familiar with the end-users' work practices. "The work on mathematical model-building and on systems analysis, in addition to its epistemological difficulties, had come up against the major stumbling block of its distance from an effective system of educational data collection and classification" (Papadopoulous 1994: 51).

Another tool that the CSTP produced in the 1960s was a 'Handbook' for education statistics, *Methods and Statistical Needs for Educational Planning* (1967). Statisticians from member states were consulted during its construction. The Handbook was used to reclassify each member state's educational system into an OECD standardized classification scheme (OECD 1981a). These were published in a series of volumes entitled *Classification of Educational Systems* from 1972 to 1975. The classification system acted as a boundary object – it helped coordinate OECD work with its member states. Classification systems embody abstract concepts that allow for generalizations to be made. At the
same time, they can be locally contextualized so that statisticians in OECD member states could fit their educational systems into the classification scheme (Bowker and Star 1999).

An important development in the OECD's organizational structure was the creation of the Centre for Educational Research and Innovation (CERI) in 1968. Initially conceived as a 3-year exploratory program, funding for CERI came from a $1 million grant from the Ford Foundation and a $750,000 grant from the Royal Dutch Shell Group of Companies (another $264,000 was set aside in support of national projects) (Austin 1972). When the future of CERI came up for review in 1971, member states voted to make it a permanent program within the OECD. In contrast to the Education Committee, CERI remained a decentralized project financed by those member states who participated in it. CERI was given its own Governing Board. Participation in CERI was composed of national experts from each participating state (OECD 1971).

The OECD Council viewed CERI as the "technical and operational wing focusing on research and development work and experimentation on strategic innovations in the educational systems" (OECD 1971: 67). Planned activities for CERI in the 1970s included "research into the relations between education and society," "innovations in the teaching/learning process," and "strengthening of national and international arrangements for educational innovation" (OECD 1971; OECD 1972). Because of its decentralized organizational set up, CERI was a venue for pursuing projects that certain member states were interested in.
In the 1980s, CERI would become the home of the OECD’s international educational indicators project.

The OECD was restructured in 1970 so that its work focused more on the inter-related aspects of economic policy. The Committee for Scientific and Technical Personnel (CSTP) was changed into the Education Committee. The Education Committee had a broader mandate than that of the CSTP reflecting the greater role the OECD assigned to education in both the economic and social spheres. Both CERI and the Education Committee were located within the Directorate for Scientific Affairs. When the OECD was reorganized in 1975, they became part of the Directorate for Social Affairs, Manpower and Education (Papadopoulos 1994, Henry et al 2001).

In response to the 1970 Paris Conference on Policies for Educational Growth, the Education Committee set up the Working Party on Educational Statistics and Indicators (OECD 1970; Papadopoulos 1994). The Working Party published in 1973 *A Framework for Educational Indicators to Guide Government Decisions*. In a follow-up report, *Indicators of Performance of Educational Systems*, an attempt was made to develop a set of educational social indicators for evaluating educational system performance. The report discussed the theoretical and technical complexity involved in undertaking such a task. The authors noted that it was difficult to measure the “total product of education” since a state has in place a set of interdependent goals for its system of education (OECD 1973: 31). Such a task would be greatly facilitated when

---

65 In 1991, the Directorate was renamed the Directorate for Education, Employment, Labour and Social Affairs (DEELSA).
educational productivity was narrowly defined in terms of measurable objectives as it was in the 1980s using New Public Management Theory. At the same time, the follow-up report did identify two areas where achievements tests can be used as "possible" indicators. Examples provided were measures for the "amount of knowledge and skills transmitted to individuals through the educational process" (38). Examples of such measures included the National Assessment for Educational Progress (NAEP) and the IEA studies. The other area was in 'Equality of Performance' where "achievement scores by race, sex and I.Q." would be used to measure progress towards "educating for equality" (78-79).

In 1974 and 1975, the OECD published Volumes I and II of the *Educational Statistics Yearbook*. The Yearbook used the Handbook, the OECD's system of educational classification, and the Framework as its foundation (OECD 1974). These publications did not identify any educational indicators or statistics of international student achievement. Student achievement was conceived as measuring school effectiveness but it was not regarded as a performance indicator and output measure as it is today. In the next section, we shall see how such changes in organizational practices would come about as the New Public Management theory took root in the 1980s. These new practices required that output indicators such as student performance be equated with measures for school performance.

OECD member states did not readily adopt the educational indicators that were developed in the 1970s. Their educational statistical systems were
equipped to measure inputs for planning purposes rather than outputs. There was not enough interest or commitment generated in changing their systems or adjusting their educational research approaches to a new set of indicators. Work on educational social indicators was put on the back-burner, to be revisited in the next decade when quality and accountability of educational systems became central concerns for OECD member states (Papadopoulos 1994).

In the late 1970s, OECD member states faced high youth unemployment rates as a large number of students were graduating from secondary and post-secondary educational institutions. The OECD created a Joint Working Party consisting of representatives from both the Education Committee and the Manpower and Social Affairs Committee. In its 1977 report, *Education and Working Life*, the Working Party linked the success of educational initiatives with corresponding employment initiatives. The Working Party also addressed the issue of vocational education and training. Since schools were viewed as implanting in students the basic competences for working life, the Working Party developed a framework for such competences. Experts were brought together to discuss these competences. They grouped them into three areas: basic skills, personal and social skills, and knowledge about working life. They pointed out that these competences would help "ease the transition from education to working life" but the experts emphasized that "[s]uch efforts cannot

---

66 In the 1990s, the OECD would later take up research in this area through the international educational indicators project's work on cross-curricular competencies and launch a study entitled the *Definition and selection of competencies: theoretical and conceptual foundations* (see Rychen & Tiana 2004).
help reduce unemployment: that is the task of manpower and employment policies" (OECD 1982: 13).

In the next section, I will discuss the shift in OECD education policy approaches that occurred in the late 1970s and 1980s.

Table 4: Education inside the OECD – Committees and Governing Boards

<table>
<thead>
<tr>
<th>Directorate</th>
<th>Year</th>
<th>Committees/Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directorate for Scientific Affairs</td>
<td>1961 – 1970</td>
<td>CSTP</td>
</tr>
<tr>
<td>1968 – 1970</td>
<td>CSTP</td>
<td>CERI (exploratory)</td>
</tr>
<tr>
<td>1970 – 1975</td>
<td>CERI Governing Board (permanent)</td>
<td></td>
</tr>
<tr>
<td>1971 – 1975</td>
<td>Programme on Educational Building (PEB)</td>
<td></td>
</tr>
<tr>
<td>1972 – 1975</td>
<td>Programme on Institutional Management in Higher Education (IMHE)</td>
<td></td>
</tr>
<tr>
<td>1972 – 1975</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directorate for Social Affairs, Manpower and Education</td>
<td>1975 – 1991</td>
<td>Education Committee</td>
</tr>
<tr>
<td>1975 – 1991</td>
<td>CERI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PEB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IMHE</td>
<td></td>
</tr>
<tr>
<td>Directorate for Education, Employment, Labour and Social Affairs</td>
<td>1991 – 2002</td>
<td>Education Committee</td>
</tr>
<tr>
<td></td>
<td>CERI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PEB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IMHE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Programme for International Student Assessment (PISA)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1997 – 2002</td>
<td></td>
</tr>
<tr>
<td>Directorate for Education</td>
<td>2002 – present</td>
<td>Education Committee</td>
</tr>
<tr>
<td></td>
<td>CERI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PEB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IMHE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PISA</td>
<td></td>
</tr>
</tbody>
</table>

Neoliberalism and Changes in Educational Policy Prescriptions

In this section, I discuss OECD education activities from the late 1970s to the 1980s. I begin by explaining the international political economic context during this era and the influence of neoliberalism on both economic and education policy. I then discuss neoliberal educational reforms in OECD member states. I
trace the history of an international system for classifying educational programs which played a significant role in facilitating relations of communication and coordination among international organizations involved in the global architecture of education.

Throughout the 1970s, OECD member states faced high rates of inflation, slow economic growth, high rates of unemployment, and large balance of payment deficits. Keynesian economic policies neither succeeded in stimulating growth nor in reducing unemployment rates. The OECD altered its policy recommendations in the 1980s to a neoliberal approach. The OECD began to recommend and support 'positive adjustment policies' that relied on market mechanisms for the allocation of labour and capital. In 1980, the Secretary-General advised member states not to "impede the adjustment process" and to "accept constraints on total welfare spending in order to provide increased investment" (OECD 1980b: 7, 9). Henceforth, the macro-economy was to be managed using a combination of fiscal and monetary policies. The following year, he emphasized the linkages that existed between trade policy and macro-economic policies, advising member states not to implement policies that "distort international competition" (OECD 1981b: 10).

Changes in economic policy approaches reflected a neoliberal rationality. The OECD, which has always promoted a free market economy, now shifted economic practices away from Keynesianism to neoliberalism. This shift was also accompanied by a new regime for governing the individual. Instead of the state being responsible for the individual, there was an emphasis on the
entrepreneurial self within a “new individualised and privatized consumer welfare economy.” Under a neoliberal rationality, “[h]uman capital theory is rejuvenated in a privatized rather than statist or public form” (Peters 2001: 60).

Neoliberal ideas were transferred to the governance of education systems. Attempts were made to turn public schools into quasi-markets where students are consumers of educational services. This competitive market model encompassed a set of policies that ranged from privatization of schools and the institution of charter schools, to changing school funding formulas and school ranking according to standardized test results. Competitive accountability required measurable outcomes. Standardization of inputs and outputs was essential to measuring and quantifying outcomes (Morgan 2006).

The role of education was increasingly viewed in instrumental terms: to reduce unemployment rates by developing the appropriate skills and competences in students for ‘working life.’ Equipped with the right skills, students entering the labour market could easily adjust to a technologically-driven and knowledge based society. OECD educational activities were concerned with the development of a creative and highly skilled competitive labour force (OECD 1983). Such practices “serve and symbolise the increasing colonisation of education policy by economic policy imperatives” (Ball 1998: 122).

In order to turn public schools into efficiently run organizations that could do more with less, new managerial practices were adopted from the private sector. These practices came to be known as the New Public Management
Theory (NPM). NPM consists of a set of ideas and practices derived from new institutional economics\textsuperscript{67} and business-type managerialism (Hambleton 2002: 156). New institutional economists view NPM as a "general theory about how government can get things done" and underline the fact that "NPM is not about politics." From this perspective, bureaucracy is viewed as inefficient because it cannot "think in terms of unit costs." Instead, bureaucracy runs itself on "administrative law and budgetary appropriations" where efficiency can never exist. In order to create efficiency, bureaucracies have to be turned into competitive organizational structures where there is "open entry and contestability." Such a structure is to be created through business-type and market-oriented managerial practices (Lane 2000: 304-306).

There were two educational agendas being pursued but in varying "intensities" among OECD member states:

Stated in more general terms, two complexly related policy agendas are discernible in all the heat and noise of reform. The first aims to tie education more closely to national economic interests, while the second involves a decoupling of education from direct state control. The first rests on a clear articulation and assertion by the state of its requirements of education, while the second gives at least the appearance of greater autonomy to educational institutions in the delivery of those requirements. The first involves a reaffirmation of the state functions of education as a 'public good', while the second subjects education to the disciplines of the market and the methods and values of business and redefines it as a competitive private good (Ball 1998: 125).

OECD member states varied in their implementation of this dual agenda of educational reforms. In social democratic states such as Denmark, Norway,
Sweden and Finland, the reforms were less intense (Ball 1998: 125). These states maintained their social democratic orientation for equality and comprehensiveness in their education systems while pursuing neoliberal reforms such as "devolving responsibilities most extensively to the municipalities and schools" (Arnesen and Lundahl 2006: 290). In corporatist regimes such as the Netherlands, the intensity of reforms was somewhat mixed (Ball 1998). For example, rather than pursuing the privatization of schools, the Dutch Ministry of Education opted for the deregulation of its educational system (Karsten 1999). High intensity reforms were instituted in liberal welfare regimes such as New Zealand and the United Kingdom. In the UK, for example, educational reforms included "parental choice, diversity of provision, competition, and decentralization of management authority" (Levin and Young 2000: 196).

The OECD, the UN and the World Bank were decisive in formulating, enhancing and spreading neoliberal policies and NPM practices in the international arena (Salskov-Iversen et al 2000). These international bodies brought experts together from various corners of the world to discuss and share best practices of this new approach to public management. NPM was valued for its ability to respond to the problems facing various states in an era of increased competition and globalization where the public sector faced mounting deficits, bureaucratic inefficiencies, and a citizenry dissatisfied with service delivery (Aucoin 1995; Shields and Evans 1998). In effect, NPM became "the transmission belt for neoliberalism" (Shields and Evans 1998: 71).
The measurement and monitoring of quality was an NPM technique borrowed from the private sector and adopted to ensure bureaucratic educational systems became competitive and efficient organizations (Ball 1998). The quality of basic education became central to OECD work and helped propel research into educational indicators (OECD 1984). An international conference on this topic was held in Washington, D.C. in May 1984. Participants at the meeting shared information on how they defined quality and the factors they used to evaluate educational outcomes (OECD 1984). Quality in Basic Education was also an important theme at a meeting of ministers of education held in Paris in November 1984. During this period, several ministries of education and government agencies published reports on the topic of educational indicators. For example, the U.S. National Center for Educational Statistics (NCES) published in 1985, *Indicators of Education, Status and Trends*; the Quebec Government published in 1988, *Educational Indicators for the Elementary and Secondary Levels*; and the Dutch Central Bureau of Statistics published in the same year *Educational Indicators for the Basisvorming* (van Harpen 1992).

In the academic arena, a growing number of educational researchers were writing on the topic of educational indicators. An important publication that consolidated and succinctly presented this knowledge was issued by the Center for Policy Research in 1986. Written by Jeanne Oakes, *Educational Indicators, a guide for policymakers* was “designed to help education decision-makers understand the legitimate roles indicators play in monitoring the condition of the
education system" (Oakes 1986:v). Oakes' publication was to be used to help construct the conceptual framework underpinning the OECD educational indicators' project (see Nuttall 1992).

In 1987 a joint conference was organized by the OECD and the U.S. Department of Education that laid out the foundation for a new project on educational indicators. In the next chapter, I chart how domestic politics led American officials to become preoccupied with compiling international educational indicators and with influencing the creation of such a project under the auspices of the OECD.

Policy concerns relating to improving the quality of basic education helped sway member states towards a more evaluative and statistical concept of education. As we noted earlier in the chapter, substantial work in the area of educational statistics had already been done by the OECD. The Working Party on Educational Statistics and Indicators had published *Educational Statistics Yearbook* in 1974 and 1975. It revised the Yearbook and published it in 1981 as the *Educational Statistics of OECD Countries*, recommending that OECD educational statistics be harmonised with those of UNESCO and EUROSTAT.68 In the early 1980s, the OECD adopted UNESCO's educational classification system, the International Standard Classification of Education Systems (ISCED) (OECD 1981).

---

68 Formerly known as the Statistical Office of the European Communities, it was renamed Eurostat in 1973 for practical linguistic reasons - for the "harmonization of publications in the wake of enlargement" which now included Denmark, Ireland and the United Kingdom (de Michellis and Chantraine 2003: 69).
I want to discuss briefly the history of ISCED because, as a classification system, it was crucial in facilitating relations of communication and coordinating work among major organizational players in the global architecture of education.

The International Standard Classification of Education Systems

The OECD opted to create its own "pragmatic" (OECD 1981a: 206) classification system that addressed the needs of its member states even though the International Recommendation on Educational Statistics, a minimal standard for reporting educational statistics, was adopted by UNESCO in 1958 (Fernig 1963). This standard was formalized and expanded upon with the creation of UNESCO's own educational classification system, the International Standard Classification of Education (ISCED).

Implemented in 1976, ISCED was conceived as "an instrument suitable for assembling, compiling and presenting statistics of education both within individual countries and internationally" (UNESCO 1976: 1). Its purpose was to facilitate the compilation of comparative educational statistics. As has been the case since 1958, ISCED classified educational programs into different levels. Under ISCED, however, a provider of educational programs need not be limited to an educational institution. ISCED classified educational levels according to the "minimum prior education required to undertake the program" (Goldstone 1973: 395). Thus, ISCED's scheme accounted for the cross-cultural variations that existed between states in their delivery of educational programs.

Classification systems do the same work as boundary objects (Bowker and Star 1999). Boundary objects help to coordinate work by being "weakly..."
structured in common use" and "strongly structured in individual-site use" (Star and Griesemer 1989: 393). ISCED operated as a boundary object. It was a classification system that could be customized to address the categorization needs at the local level. ISCED drew on abstract and generalized definitions to accommodate the diversity that existed in the delivery and organization of educational programs (i.e. weakly structured in common use). In this way, ISCED categories can be "tailored to meet the needs of any one community" (Bowker and Star 1999: 16). When it came to implementing ISCED at the national level, then ISCED was used for more specific and concrete categorical work (i.e. strongly structured in individual-site use). The fitting together of objects with categories is a strongly structured process that is imposed by ISCED at the local level. Work has to be done to fit national educational programs and educational delivery system into ISCED categories.

ISCED 1976 was constructed to allow for ambiguity. A residual category was created especially for those objects that could not be fitted with the classification system. In the next chapter, I will discuss how the OECD came to adopt ISCED and how, in 1997, ISCED was revised precisely because it was inadequate for constructing categories for an international system of educational indicators. ISCED as a boundary object was crucial to facilitating relations of communication across communities of practice and in coordinating work across multilateral and regional organizations. By using a classificatory system such as ISCED, cooperative work in the area of educational statistics can be more easily undertaken by various agents with the global architecture of education. More
specifically, organizations such as the OECD, UNESCO and EUROSTAT collaborated to ensure they have adequate reporting and monitoring mechanisms to achieve their educational objectives.

**Figure 3: ISCED Classification of Educational Programs (1976)**

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Education preceding the first level</td>
</tr>
<tr>
<td>1</td>
<td>Education at the first level</td>
</tr>
<tr>
<td>2</td>
<td>Education at the second level, first stage</td>
</tr>
<tr>
<td>3</td>
<td>Education at the second level, second stage</td>
</tr>
<tr>
<td>5</td>
<td>Education at the third level, first stage, of the type that leads to an award not equivalent to a first university degree</td>
</tr>
<tr>
<td>6</td>
<td>Education at the third level, first stage, of the type that leads to a first university degree or equivalent</td>
</tr>
<tr>
<td>7</td>
<td>Education at the third level, second stage, of the type that leads to a postgraduate university degree or equivalent</td>
</tr>
<tr>
<td>9</td>
<td>Education not definable by level</td>
</tr>
</tbody>
</table>


UNESCO's compiled educational statistics were based on information provided by three questionnaires that were issued to member states. ISCED 1976 was used as the organizing scheme for defining educational systems and programs. The first questionnaire focused on pre-primary, primary and secondary education; the second on educational expenditure and financing, and the third, on higher education. OECD and UNESCO worked on adapting these questionnaires to address their informational needs but found that two sets of separate questionnaires would be more useful – one for the OECD member states and one for UNESCO member states excluding OECD states. Based on the replies received from these questionnaires, the OECD published *Education in OECD Countries: a compendium of statistical information* annually from 1986 to 1989.
Conclusion

In this chapter, I examined how the OECD's organizational units built the technical capacity for measuring educational inputs and outputs. As the OECD developed its technical expertise, it also established relations of communication and an infrastructure for coordinating work across communities of practice. The OECD work in this area was linked with other educational measurement activities being undertaken by the European Union, the World Bank and the United Nations. Increasingly, OECD educational objectives were becoming integrated into a global architecture of education.

We saw how OECD educational activities were initially concerned with developing a set of tools for educational planning purposes as states invested in expanding their educational systems. These quantitative methods for educational measurement developed by the mathematicians were too complex for educational policymakers, planners and administrators. Even though the educational indicators that the social scientists created were too general to be used by ministries of education, the OECD did succeed in developing an educational statistical handbook that assisted in reporting educational indicators. During the neoliberal era, the OECD played an important role in the dissemination New Public Management ideas and in instilling these ideas in policy practices. Educational policy directions turned away from investment in educational capacity to monitoring quality and effectiveness of educational systems.
In the next chapter, I examine an adjustment that took place in the power bloc of international educational measurement. Two major developments set a new trajectory for the field of international educational measurement.
Chapter 6
American Influence and the Adjustment of the Power Bloc

Introduction

The two previous chapters discussed developments at the IEA and the OECD. The formation of a power bloc in international educational measurement took shape as communities of practice built technical capacity, relations of communication and power relations. In this chapter, I argue that an adjustment took place in the power bloc that resulted in a new trajectory in the field of international educational measurement. The evidence suggests that this trajectory was influenced by domestic American politics and American influence in the international arena under the Reagan (1981-1989) and George H. W. Bush (1989-1993) Administrations.

Chapter 6 details American domestic developments that led officials with the U.S. Department of Education to use multilateral channels as an instrument for achieving their policy objectives. It explains how American officials achieved consensus within the OECD for an evaluative and technical approach to education, and describes the manner in which the OECD and the IEA were enrolled into an American problematization of educational indicators. It traces the creation of an international educational indicators infrastructure that came to be known as the INES Project. In the final section, I introduce a new community of practice which was formed to assess adult literacy levels. The use of 'literacy' as a measure was crucial to how the OECD would objectify the skills and competencies of the future worker in a knowledge-based economy.
The impact of this new trajectory has been the dissemination of an American model of education. The American model is guided by accountability frameworks that monitor performance through the use of educational indicators. The model also uses student achievement on national and international tests to assess the output of educational systems.

**A Nation at Risk**

To understand how the American model for educational accountability became entrenched internationally, we need to examine the context of American domestic politics. In 1981, Terrell Bell, the Reagan Administration’s Secretary of Education created the National Commission on Excellence in Education to report on the quality of the American educational system. The Commission’s report, *A Nation at Risk: the Imperative for Educational Reform*, was published in 1983 and instigated the widespread belief in an educational “crisis” (Berliner and Biddle 1995:139). Using war metaphors, the report pointed to the vulnerability of American competitiveness caused by its mediocre education system.

If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen to ourselves. We have even squandered the gains in student achievement made in the wake of the Sputnik challenge. Moreover, we have dismantled essential support systems which helped make those gains possible. We have, in effect, been committing an act of unthinking, unilateral educational disarmament (National Commission on Excellence in Education 1983).

69 Berliner & Biddle state that this “incendiary document” was based on “false evidence” (1995:3). Berliner & Biddle (1995) present evidence that the education system was not in a crisis as *A Nation at Risk* claimed.
A Nation at Risk used a series of 'risk' indictors as factual representation of the poor performance of American students on national and international tests. A concern of the Department of Education was to be able to track the performance of American schools both nationally and internationally.

At the national level, the Department of Education had in place since 1969 the National Assessment of Educational Progress (NAEP). NAEP tested American students on their achievement in various subjects, including reading, mathematics and science. In the 1990s, a 'state' NAEP was also introduced, providing the Department of Education with state-by-state comparisons of educational achievement. Two individuals who played a crucial role in transforming NAEP into a state-by-state assessment were later to become important participants in the field of international testing: Eugene Owen, who became Chair of the OECD's project on international educational indicators, and Ina Mullis, who became the co-director of the IEA International Study Center located in Boston College (Interview Respondent #6, 29 August 2006).

American officials with the Department of Education argued that they wanted reliable international educational indicators. Since the Americans had withdrawn from UNESCO in 1984, American policymakers looked to the OECD as a venue for gathering reliable international indicators. In fact, the OECD would provide a policy forum where delegates from rich countries could have

---

70 NAEP test sample student populations aged 9, 13, and 17 and, since the 1980s, in grades 4, 8, and 12.
closed-door meetings without the politicized environment that characterizes United Nations agencies (Salzman and Terracino 2006).

International student achievement results were desirable so as to track the relative performance of American students. This became an imperative issue in 1990 when President George H. W. Bush and the Governors agreed on six National Education Goals. One of these goals was for American students to be "first in the world in mathematics and science education" by the year 2000. American policymakers needed to measure student achievement regularly, reliably and consistently in science and mathematics (Salganik and Calsyn 2001). Even though the Department of Education had funded international evaluation studies since the inception of the International Association for the Evaluation of Educational Achievement (IEA) in the early 1960s, it now pursued a more aggressive strategy for obtaining international comparative data on student achievement.

*A Nation at Risk* (1983) had cited the dismal results of American students on international assessments and it noted that "[i]nternational comparisons of student achievement, completed a decade ago, reveal that on 19 academic tests American students were never first or second and, in comparison with other industrialized nations, were last seven times (National Commission on Excellence in Education 1983). Data from the IEA studies were used as evidence to justify American educational reforms, but officials with the Department of Education questioned the quality of the data. The reporting of international assessment data was criticized for not providing sufficient evidence
that explained the results and for not discussing the limitations of these studies (Chabott and Elliott 2003). There was criticism of the unreliability and bias in international education data and of the narrowness of scope of what was being measured (Heyneman 1993). Furthermore, international assessments that were implemented at the national level were not funded adequately and were usually conducted by under-resourced universities (Interview Respondent #6, 9 June 2006). In order to monitor American involvement in international assessments and to provide expertise in this area, the federal government created the Board on International Comparative Studies in Education (BICSE) in 1989.

Senior government administration officials backed the Department of Education’s strategy politically and financially. They believed that gathering “timely and accurate information” was the “most powerful lever for improving education” (Finn 2000). The strategy for gathering educational information was implemented by the National Center for Education Statistics (NCES).

The IEA as a Venue for an International Assessment

Developing reliable, international student assessments was one of the components of the NCES strategy for gathering international educational indicators. The U.S. Educational Testing Service (ETS) offered to model the international test on the National Assessment of Educational Progress (NAEP). ETS, a non-profit organization and a leader in the area of testing, had been administering the NAEP since 1983 (Vinovskis 1998). In 1988, the Department of Education and the National Science Foundation funded an international
version of the NAEP known as the International Assessment of Educational Progress (IAEP) that tested students' science and mathematics achievement (BICSE 1995). Only five countries and four Canadian provinces participated in IAEP-I.\(^1\) According to one account, this event marked an end to the IEA "monopoly" over the international assessment arena, since ETS was responsible for the IAEP's creation and analysis of the data (Bottani and Vrignaud 2005: 16).

A second IAEP or IAEP-II was administered in 1991. Nine and thirteen year-old students from twenty national educational systems participated in this second survey. Unfortunately for American officials, the IAEP was criticized for its American bias since it was modelled on the National Assessment for Educational Progress. IAEP was therefore discontinued. The National Center for Education Statistics (NCES) was once again left without an international student assessment that would serve as a measurement tool for assessing progress "towards the standards set by the highest-performing students in the world" (Owen, Stephens, Moskowitz, and Gil 2004: 8).

The other venue for the administration of an international student assessment was the IEA. Until the 1980s, the IEA had been conducting its mathematics and science achievement tests separately. The National Center for Education Statistics (NCES) developed a proposal for the design of a recurring mathematics and science assessment (Interview Respondent #12, 15 August 2006). The Americans approached the IEA in the early 1990s with this proposal and requested that it expand its studies to include both science and

\(^{1}\) 24,000 13-year-olds in Ireland, Korea, Spain, the United Kingdom, the United States, British Columbia, New Brunswick, Ontario and Quebec were tested in the first IAEP or IAEP-I.
mathematics in one assessment (Interview Respondent #9, 29 June 2006; #12, 15 August 2006), and the IEA accepted the American proposal in 1990. The test came to be known as the Third International Mathematics and Science Study or TIMSS. In 1991, NCES began funding another project entitled the Survey of Mathematics and Science Opportunities (SMSO) that was to provide the theoretical foundation for the selection of test items. The SMSO project's director was Bill Schmidt and the project was located at Michigan State University (Interview Respondent #15, 29 August 2006; BICSE 1995).

TIMSS was the "largest and most ambitious international comparative study of student achievement to date" (Martin 1996:1). Students from more than forty educational systems and five grade levels were tested. The organizational structure of TIMSS depended on the collaboration of researchers working in national research centres around the world. The International Study Center at Boston College was responsible for the overall management of the study, including supervision of its design and implementation, data analysis of the results, reporting international results and quality assurance. Various TIMSS functions were shared among research centres. For example, Statistics Canada advised National Research Coordinators on sampling related tasks and compilation of samples; the Australian Council for Educational Research (ACER) assisted in developing achievement tests, in conducting psychometric analyses of field trial data, and in the development of scaling software and in scaling the achievement test data. Westat Inc., an employee-owned research
organization, worked with Statistics Canada and the NRCs to ensure that sampling plans met the TIMSS standards (Gonzalez, Smith and Sibbers 1998).

TIMSS was considered to be a groundbreaking international assessment. One of the respondents identified one particular area where TIMSS contributed to advancing the science of international student measurement. TIMSS incorporated both multiple choice and open response items or free-response items. Through a collaborative process, a new coding practice for free-response items was created. Norwegian curriculum and statistical experts constructed a new method for scoring these types of open responses that came to be known as two digit coding. “The first digit, ranging between 1 and 3, would be used for a correctness score, and the second digit would relate to the approach used by the student. Numbers between 70 and 79 would be assigned for different categories of incorrect response attempts, while 90 would be used if the student did not respond” (Lie, Taylor and Harmon 1996:7-5). This was described as a kind of revolution in international student testing (Interview Respondent #15, 29 August 2006). The coding of open-response items enhanced the reliability of students’ answers on these types of questions by providing a standardized technique to be used by all scorers. It reduced the ambiguity associated with constructed student responses.

TIMSS was very costly and was riddled initially with organizational problems. In fact, these problems nearly threatened the success of the TIMSS. NCES had invested a considerable amount of money - nearly $42 million - in a very costly endeavour (Interview Respondent #18, 22 September 2006). The
management of TIMSS was reorganized in 1993 after a crisis meeting. At this meeting, NCES was influential in deciding how the study was to be managed (Interview Respondent #14, August 27, 2006). The main study centre was moved from the International Coordinating Center at the University of British Columbia to Boston College’s International Study Center. The new project director became Al Beaton replacing David Robitaille (Martin and Kelly 1997). With TIMSS under new management, the U.S. Department of Education ensured that the study had a better chance of success.

In 1989, the IEA was reorganized and relocated to Amsterdam. American government officials were able to oversee and monitor the IEA when Gordon Ambach became an IEA Board member. As the IEA website notes, Ambach “played a key role in transforming IEA into a professional organization” and “contributed significantly to the creation of an IEA-policy for the cycle of studies and to the development of links with funders” (IEA 2007).

Ambach reported on the IEA to the Board on International Comparative Studies in Education (BICSE) (Interview Respondent #12, August 15, 2006). As noted earlier, BICSE was created in 1988 to advise the federal government on large scale, cross-national education surveys. Under the Clinton Administration, BICSE’s mandate was expanded in 1998 to include other kinds of comparative international studies in education, to oversee American involvement in international assessments and to provide expertise in this area (Viadero 2003).72

---

72 BICSE was ended in 2003 by George W. Bush.
Having discussed the enrolment of the IEA into an American problematization of an international student assessment, I will now turn to how the Americans enrolled the OECD into their problem structure for the production of international educational indicators.

The OECD as a Venue for the International Indicators Project
After the publication of *A Nation at Risk* in 1983, the U.S. Department of Education “pressed, powerfully and persistently, for quality to become a top priority for the Organisation’s (OECD) work” in education (Papadopoulos 1994:181). Quality was defined in terms of learning outcomes as opposed to more encompassing notions of education’s contribution to society. Along with a shift towards making quality central to OECD education policy research, the American delegates to the OECD “pressure[d]” the OECD to develop comparative educational output and input indicators (Heyneman 1993:375). They even threatened to “withdraw support from CERI” if their demands were not addressed (Henry *et al* 2001:87).

The analysts at the Centre for Educational Research and Innovation (CERI) hesitantly carried out the work on international educational comparative indicators for the OECD. The majority were researchers linked with the larger teacher trade union movement (Interview Respondent #8, 19 June 2006). Most CERI staff thought it was “unprofessional to try and quantify such indicators, that it would oversimplify and misrepresent OECD educational systems” (Heyneman 1993:375). They were very “suspicious” of American attempts at developing
such educational indicators and thought they were politically motivated by a
neoconservative presidency. The U.S. Department of Education felt it was
important to create the conditions inside CERI for the capacity to evaluate and
measure educational outcomes (Interview Respondent #8, 19 June 2006).

Officials with the National Center for Education Statistics (NCES)
believed it was in America's national interest to compare itself against other
industrialized states but it had to do so according to a common framework that
contained clearly defined indicators (Interview Respondent #6, 9 June 2006).
The U.S. Department of Education decided that the OECD was an appropriate
venue for the implementation of such a project. Officials from NCES took the
lead in funding and building an infrastructure for the production of reliable and
comparable international educational indicators.

**International Indicators and Evaluation of Educational Systems (INES)**

**Project**

The INES Project brought together educational researchers who work inside
government agencies, universities and research centres. The two key founders
of the INES Project, the OECD and the National Center for Education Statistics
(NCES), laid a plan for the collection of indicators through a progressive series
of phases: an exploratory phase, an indicators development phase, and a
production phase. It was envisaged that once these phases were completed,
the project would become part of the OECD's regular program of work.

*Phase 1: Exploratory Phase*
The topic of comparative indicators was "cautiously" explored during two meetings – the first, sponsored by the U.S. Department of Education, was held in Washington, D.C. in 1987 and the second, sponsored by the French Ministry of Education, took place in Poitiers, France in 1988 (Papadopoulos 1994; Henry et al. 2001). French policymakers were allied with their American counterparts in their pursuit for international educational indicators. As Henry et al. (2001) point out, their interests converged: the French Ministry of Education was driven by its "French republican ideal" and the U.S. Department of Education by its New Public Management theory (99).

After these meetings, a "new experimental project" was initiated within CERI. In May 1988, the International Indicators and Evaluation of Educational Systems (INES) project was established "on which certain Member countries have agreed to work together in the development of a set of international educational indicators covering, in particular, student flows, student outcomes and costs and resources" (OECD 1988: 59). Ronald Gass, Director of Education with the Directorate for Social Affairs, Manpower and Education, asked Norberto Bottani to head the project. Bottani was one of the few analysts within CERI who supported the development of key educational indicators. The INES Project was established as a decentralized project. It was funded by participating states and not centrally funded by the OECD. Such a decentralized status enabled American officials with NCES to maintain influence over the direction of the project (Interview Respondent #8, 19 June 2006).
Advisory Group assisted by the OECD Secretariat oversaw the project during this phase (Henry et al 2001).

The Project was organized into five international Networks, each led by an OECD member state. The lead state “coordinated the contributions of members, supported the theoretical and technical work required, and produced the network report” (BICSE 1995). It also was responsible for funding meeting costs, communication costs, staff support and publications. Within the OECD, the Scientific Advisory Group guided the Networks’ work (BICSE 1995). The Networks were responsible for researching the feasibility and the construction of the following groupings of indicators:

- **Network 1**: Indicators on student flows
- **Network 2**: Indicators on student achievement
- **Network 3**: Indicators on school processes
- **Network 4**: Indicators on costs and resources
- **Network 5**: Indicators on attitudes and expectations towards educational Systems (Henry et al 2001:88)

The U.S. Department of Education was responsible for Network 2, Indicators on student achievement. This Network would later be renamed Network A and would become responsible for developing the PISA.

Bottani organized a general assembly for the INES Project in order to begin exploring the production of educational indicators. Held in Austria in September 1989, the general assembly brought together all five working groups to tackle the problem of finding comparable educational data. The exploratory work of the Networks was completed in December 1989 “with a surprising
degree of consensus and enthusiasm.” According to the OECD Secretariat, the indicators “received a splendid political reception” (Henry et al. 2001:88).

A restructuring of the Networks occurred at the end of 1989. Network 2 was divided into two Networks – Network A, Student Achievement Outcomes led by the U.S. Department of Education (NCES) and Network B, Education and Labour Market Destinations led by the Swedish Ministry of Education. The Chair of Network A was Dr. Gary Phillips, an expert in large scale assessments and surveys. Phillips at the time was also working on an international student assessment proposal to be presented to the IEA that combined mathematics and science achievement which came to be known as TIMSS (Interview Respondent #12, 15 August 2006). Phillips was replaced in 1994 by Eugene Owen, a sociologist with statistical training (BICSE 1995). As mentioned earlier, Owen had played an influential role in the creation of the state National Assessment of Educational Progress (NAEP).

**Phase 2: Developing Indicators**

In 1990 and 1991, members of the INES Project worked on developing and constructing a set of educational indicators. Two technical groups were organized to address issues related to data gathering and definitions. The end of the second phase of the INES Project was demarcated by an INES General Assembly meeting in Lugano, Switzerland where the first draft of *Education at a Glance* (EAG), containing data on thirty indicators was presented. During this meeting, the OECD committed itself to the publication of these indicators within

---

70 In 1995, Network A’s name was changed to Education Outcomes (BICSE 1995)
a period of one year (Interview Respondent #8, 19 June 2006; OECD 1995b; Henry et al. 2001).

The OECD’s Centre for Educational Research and Innovation (CERI), where the INES Project was based, lacked expertise and resources to embark on a project for developing educational indicators. The National Center for Education Statistics (NCES) provided financial assistance for the project. Bottani hired Albert Tuijnman as Principal Administrator to help with the production of the first *Education at a Glance* (EAG). Tuijnman’s doctoral research examined the labour market effects of adult education and training programs. Tuijnman would later become involved in the creation of a new community of practice for assessing adult literacy internationally. Additional technical expertise was obtained on a temporary basis from individuals associated with the National Center for Education Statistics (NCES) (Interview Respondent #8, 19 June 2006).

Published in 1992, the first *Education at a Glance* (EAG) was cautious in putting forth comparative educational indicators. It specified that some indicators were “provisional” since they were “considered experimental, not only because of disagreement as to their validity and applicability, but also because they were measured using data that were not specifically collected for this purpose” (OECD 1992a:12). The charts containing provisional or experimental data were coloured pink. The Learning Outcomes indicators, comprising

---

74 This information is obtained from Albert Tuijnman’s on-line CV where he is Professor of Comparative and International Education at the University of Nottingham’s UNESCO Centre for Comparative Education Research (http://www.nottingham.ac.uk/education/centres/uccer/centre-staff.phtml?name=tuijnman-albert)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
mathematics achievement, school differences in achievement and student differences in achievement, were all presented as experimental. The data were derived from studies conducted by the International Association for the Evaluation of Educational Achievement (IEA) and from the first and second International Assessment of Educational Progress. The publication of the first EAG marked the transition to the next phase of the project for international educational indicators.

Phase 3: Production Phase

The publication of two CERI reports in 1992, *Education at a Glance: OECD Indicators* and *The OECD International Education Indicators: A Framework for Analysis*, moved the INES Project to Phase 3, the Production Phase from 1992 to 1996. These developments were accompanied by an internal restructuring of the OECD's educational work. CERI's Scientific Advisory Group, which was created to help guide the work of the INES Project during its exploratory phase, was reconstituted as the Policy Review and Advisory Group (PRAG), chaired by the chairman of the CERI governing board (BICSE 1995). PRAG was set up to help "anchor" the indicators' work in policy (Henry et al. 2001:89). The status of educational indicators work within the OECD was enhanced by this reorganization.

Within the OECD's Directorate for Education, Employment, Labour and Social Affairs (DEELSA), a Statistical and Indicators Division was created that combined the resources of both CERI and the Education Committee. Henry et al. (2001) interpret the restructuring as symbolizing the "mainstreaming of the
OECD's indicators work in education, from developmental status to being a core activity of the Organisation" (89). This mainstreaming of educational statistical work was also consistent with the OECD's broader commitment to the compilation of statistical information as it had established the new Statistics Directorate in 1992. The OECD's Chief Statistician's mandate was to "coordinate and rationalise statistical activities within the OECD as a whole" and to "promote more efficient use of the collective statistical resources" regionally and internationally (OECD 1994:17).

The organizational changes entailed changes in practices. OECD hired new employees with statistical and technical backgrounds to do its education work. For example, Andreas Schleicher, who had project management experience working for the IEA, was hired in 1994. Schleicher would eventually become head of the OECD's Indicators and Analysis Division within the Education Directorate, responsible for both the INES Project and for the Programme for International Student Assessment. At the national level, OECD member states committed resources and assigned national coordinators to collect INES Project data. This contributed to the "diffusion of an indicator culture within education circles" (OECD/CERI, quoted in Henry et al 2001:89).

In preparation for establishing the INES Project as a permanent component of OECD work, the Third General Assembly was organised in Lahti, Finland in June 1995. It was attended by over two hundred delegates representing twenty-three member and ten non-member countries. The delegates came together "to assess the work INES had carried out since 1992,
to list unfinished activities, and to identify future priorities" (OECD 1995: 88). The delegates decided that the INES Project should focus on specific domains involving fewer indicators. Furthermore, they stipulated the characteristics of these indicators: "Indicators must be policy relevant, policy malleable, and able to reflect changes in important aspects of education systems; they also should be reliable, valid, and comparable. Indicators should be presented in an accessible manner" (OECD/INES 1995b: 2).

At the Third General Assembly, two initiatives were introduced that were significant to the infrastructural work the OECD was doing in the area of educational indicators. The first was the introduction of a new data strategy for collecting indicators on student outcomes. *A Strategy for Producing Student Achievement on a Regular Basis* encompassed the foundational elements for an OECD international student assessment. In Chapter 7, the Data Strategy will be discussed in greater detail. The second initiative was a recommendation for the OECD to contribute to the revision of the International Standard Classification of Education (ISCED) that was being undertaken by UNESCO. The INES Project faced problems developing useful indicators due to the lack of "conceptually adequate and operational definition of levels of education." More specifically, difficulties were encountered developing comparable indicators on "costs and resources, graduation rates and level of educational attainment in the population and indicators on labour market outcomes" (OECD 1999b: 7).

A Technical Group consisting of twenty to thirty advisors was established as part of the INES Project. This group played a pivotal role in implementing the
newly revised ISCED, known as ISCED 1997, in OECD member states. It published an implementation manual mapping each OECD member state’s educational system to the ISCED 1997 classification. Even though ISCED operated as a boundary object allowing for flexibility in implementation, it was operationalized as a standardized package for OECD member states. By standardized package, I mean that it contained both flexible concepts as well as standardized methods. ISCED 1997 allowed for some definitional flexibility as in ‘levels of education’. At the same time, OECD member states implementing ISCED had to follow very “specific instructions” to ensure comparability. Their educational systems were mapped to the ISCED 1997 scheme (OECD 1999b: 8). The Technical Group monitored the implementation of ISCED 1997 by OECD member states.

**Figure 4: ISCED Classification of Educational Programs (1997)**

| Level 0 | Pre-primary education |
| Level 1 | Primary education or first stage of basic education |
| Level 2 | Lower secondary or second stage of basic education |
| Level 3 | (Upper) secondary education |
| Level 4 | Post-secondary non-tertiary education |
| Level 5 | First stage of tertiary education |
| Level 6 | Second stage of tertiary education |


ISCED 1997 allowed for less ambiguity in categorical work. By categorical work, I mean “the work of attaching things to categories and the ways in which those categories are ordered into systems” (Bowker and Star 1999: 286). Any residual categories that existed in ISCED 1976 were removed to ensure that all educational levels were fitted to the ISCED classification.
scheme. The original ISCED of 1976 had a residual category '9' for 'education not definable by category'. This category was removed in ISCED 1997.

Within each level of education, there was a residual category of '89' for other fields and a residual category '99' for other groups of programmes within each field. In ISCED 1997, these residual categories were also removed. Those doing categorical work did not have any leeway to report anything as residual. As ISCED 1997 clearly indicates:

**Not known or unspecified**
(This category is not part of the classification itself but in data collection '99' is needed for 'fields of education not known or unspecified'.)
(UNESCO 2006: 45)

Not only was there less ambiguity with a revised classificatory system in place, but also the INES Project was reorganized so that its decision making structure was less inclusive and more efficient. The INES Project’s structure was reconfigured in order to enable it to achieve the objectives set out by the Lahti General Assembly. The Policy Review and Advisory Group (PRAG) was changed to a Steering Group responsible for “overall policy direction” (OECD/INES 1996b: 2). This group was chaired by Tom Alexander, the Director of the Directorate for Education, Employment, Labour and Social Affairs (DEELSA) (OECD/INES 1997a). The role of National Coordinators was clearly defined in terms of advising the Steering Group and assisting with dissemination. To ensure the Networks were productive, they would be given specific goals to be achieved within a given time period (OECD/INES 1996b: 2).
Until 1996, the INES Project operated as a vast network consisting of three hundred to five hundred people worldwide (Interview Respondent #8, 19 June 2006). A consultative process was in place. Groups from various disciplines had discussions and developed recommendations that were then presented to CERI’s Policy Review and Advisory Group (PRAG) that was responsible for overseeing the INES Project. Bottani (1999) described the process as a “bottom-up” approach that “mobilize[d] a considerable amount of resources and energy towards a common shared objective.” It was a democratic but very time consuming and lengthy process that conflicted with the needs of member states. Member states wanted a stream-lined process with shorter timelines (Interview Respondent #8, 19 June 2006). In 1997, this consultative process was “dismantled” and replaced with a “top-down, authoritarian” approach (Bottani 1999).

The INES Project was successfully executed as ambiguity was reduced both in the way indicators were to be classified using ISCED 1997 and in the manner in which the INES Project was organized. Its infrastructure and technical capacity were built methodically using several coordinating mechanisms. The Network structure helped organize the tasks in each area of indicator work. The direct involvement of educational researchers and statisticians in this work ensured that the tools that were created addressed each educational system's requirements. Because it was a well-funded project, INES was financially stable. Its success was demonstrated by the popularity of *Education at a Glance* which the OECD described as its “flagship publication”
The final structure of the INES Project consisted of three networks, each with a lead state, and a Technical Group of experts:

Network A: Indicators on learning outcomes (U.S. Department of Education)
Network B: Indicators of the social and economic outcomes of education\(^7\) (Swedish Ministry of Education, Research and Culture)
Network C: Indicators on structures and processes of schools (Dutch Ministry of Education, Culture and Science)

Technical Group

The INES Project remained a decentralized project and was never integrated into the OECD as a centralized program. According to one of the respondents, the American, British and the Canadian educational policymakers wanted the INES Project to remain autonomous from the OECD Secretariat. The decentralized status allowed member states control over the direction of the INES Project. If member states were to lose interest in the INES Project, then the OECD Secretariat would likely turn it into a “classic statistical collection agency” such as UNESCO’s Institute for Statistics (Interview Respondent #6, August 29, 2006). The OECD Secretariat did make one more change to the governance structure of the INES Project: it created a joint session of both the CERI Governing Board and the Education Committee. This joint session of governing board members oversaw INES Project work and ensured that the INES Project appeared on the “OECD radar” (Interview Respondent #19, September 24, 2006).

\(^7\) Representatives from Eurostat also participate in Network B.
As OECD member states pursued the development of educational indicators, another policy area came to the forefront: adult functional literacy. I will discuss how an international assessment for adult literacy skills was created.

The International Adult Literacy Survey

In this section, I trace the creation of a community of practice that measured adult literacy skills internationally. Members of this community created the International Adult Literacy Survey (IALS). The tools and artefacts they developed were later deployed to help construct an OECD-based student assessment. The technical capacity built through the construction of an instrument for measuring adult functional literacy provided a direct measure for assessing the level of human capital formation. Up to this point, statisticians had relied on educational attainment such as the grade level reached as a proxy for adult functional literacy (OECD 1992b).

The IALS community members were initially concerned with measuring adult literacy skills nationally. American and Canadian educational researchers, psychometricians and statisticians executed the first adult literacy assessments in the 1980s as educational researchers developed direct assessment methods for measuring adult literacy levels. The U.S. Department of Education contracted the Educational Testing Service (ETS) to undertake a literacy survey. The Young Adult Literacy Survey (YALS), which was administered in 1984, was considered by educational researchers to be a groundbreaking innovation in quantitatively defining and measuring literacy. The YALS applied three
distinctive scales to the tasks it asked young adults to complete — these scales included (1) Prose literacy tasks (2) Document literacy tasks and (3) Quantitative literacy tasks (Kirsch and Mosenthal 1990). The YALS was followed by a significantly larger and more comprehensive literacy survey, the National Adult Literacy Survey in 1992 which drew on the same methodology as the YALS (Sum, Kirsch and Taggart 2002).

Influenced by the American developments in this area, the Canadian National Literacy Secretariat\(^{76}\) commissioned Statistics Canada in 1989 to conduct a Canadian national literacy survey. The survey applied the same methodology as the YALS but it also measured literacy skills in both French and English. The data from the Canadian literacy survey were stable cross-culturally demonstrating the feasibility of a comparative literacy assessment (Interview Respondent #7, 9 June 2006). Paul Bélanger, Director of UNESCO’s Institute for Education, became interested in the study and invited several researchers to a meeting to discuss the possibility of undertaking an international literacy assessment. Three of these individuals were Irwin Kirsch, the Project Director of the YALS at the Educational Testing Service (ETS),\(^{77}\) Scott Murray, who led the project at Statistics Canada and Albert Tuijnman, who was from the OECD’s INES Project. During this meeting, it was decided that a larger comparative literacy study was worth pursuing (Murray, Kirsch, and Jenkins 1997).

\(^{76}\) “Founded in 1987 on a five-year mandate from Human Resources and Development Canada, the National Literacy Secretariat was intended to coordinate the provinces, the voluntary sector and the private sector for the activities of International Literacy Year” to be held in 1990 (Pound 2006: 59).

\(^{77}\) ETS had administered the YALS as part of its contract with the U.S. Department of Education.
A proposal entitled, *An International Assessment of Adult Literacy: A Proposal*, was submitted to the OECD and caught the attention of senior policymakers, particularly since there was heightened focus on the linkage between literacy and economic performance. In 1992, the OECD, in its publication, *Adult Illiteracy and Economic Performance*, had highlighted the threat low literacy rates posed to both economic performance and social cohesion (OECD 1992b; Murray *et al* 1997).

When nine states decided to undertake the first International Adult Literacy Survey (IALS) in 1994, they were “reluctant to invest large sums of money on something which had yet to demonstrate its relevance for policy” (Murray, Kirsch, and Jenkins 1997:15). Even though the OECD opted to sponsor the first IALS, Statistics Canada and the National Center for Educational Statistics (NCES) absorbed the majority of the overhead costs involved in the survey. The IALS was managed collaboratively by Statistics Canada and the Educational Testing Service (which was contracted by NCES). These agencies collected the data, analyzed it and published the results and analysis. The first IALS was considered to be very successful and subsequent rounds of IALS took place in 1996 and in 1998. A final report was issued in 2000 summarizing the results and pointing towards their policy implications.

The International Adult Literacy Survey (IALS) represented a turning point in how literacy construct came to be defined and how it came to be measured. In the IALS framework, literacy was associated with a set of skills that adults needed to have to function in the information age. Literacy skills were viewed as
<table>
<thead>
<tr>
<th>Year</th>
<th>Project/Program/Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950-1961</td>
<td>Methods for surveying the current 'stock' of scientific and technical personnel and forecasting future long-term needs</td>
</tr>
<tr>
<td>1967</td>
<td>Publication of 'handbook' for education statistics, <em>Methods and Statistical Needs for Educational Planning</em></td>
</tr>
<tr>
<td>1970</td>
<td>Education Committee creates Working Party on Educational Statistics and Indicators established.</td>
</tr>
<tr>
<td>1972-1975</td>
<td>Publication of <em>Classification of Educational Systems.</em></td>
</tr>
<tr>
<td>1981</td>
<td>Agreement reached with UNESCO on a common approach to internationally comparable statistics based on ISCED classification. OECD depends on UNESCO questionnaires for the collection of information.</td>
</tr>
<tr>
<td>1981</td>
<td>Dissolution of the Working Party on Educational Statistics and Indicators</td>
</tr>
<tr>
<td>1983-1984</td>
<td>New UNESCO/OECD questionnaires negotiated. Two sets created – one for OECD states and one for UNESCO member states not belonging to the OECD. Questionnaires are sent out and used to compile educational statistics.</td>
</tr>
<tr>
<td>1983-1986</td>
<td>Data from UNESCO/OECD questionnaires collected but not published,</td>
</tr>
<tr>
<td>1987</td>
<td>Creation of INES Project; conference on international educational indicators, Washington, D.C.</td>
</tr>
<tr>
<td>1988</td>
<td>Second conference on international educational indicators, Poitiers, France</td>
</tr>
<tr>
<td>1989</td>
<td>Publication of <em>Education in OECD Countries</em> <em>(1986/87-1989/90)</em></td>
</tr>
<tr>
<td>1994</td>
<td>First International Adult Literacy Survey</td>
</tr>
<tr>
<td>1995</td>
<td>Third General Assembly of the INES Project, Lahti, Finland (June)</td>
</tr>
<tr>
<td>1996</td>
<td>Second International Adult Literacy Survey</td>
</tr>
<tr>
<td>1997</td>
<td>ISCED 1997 negotiated with UNESCO PISA established as a regular assessment programme in the OECD. PISA Governing Board created.</td>
</tr>
<tr>
<td>1998</td>
<td>Third International Adult Literacy Survey</td>
</tr>
</tbody>
</table>
an "element of human capital" which contributes both to "personal development" and "to aggregate economic and social performance." These skills involved "information-processing" capacities in prose, document and quantitative domains (OECD and Statistics Canada 2000: 61).

More significantly, the IALS constructed a new technique for quantifying human capital. Instead of relying on educational attainment as a measure for human capital formation, workers could be directly assessed for their levels of functional literacy. Finally, the IALS required the coordination of several agents within the global architecture of education that included the OECD, EUROSTAT and the UNESCO Institute for Education.

Conclusion

The chapter showed how American domestic politics under the Republican-led Administrations of Ronald Reagan and George H. W. Bush influenced the trajectory of the power bloc. I discussed how the U.S. Department of Education enrolled the IEA and the OECD into a new problematization for international educational indicators that addressed their domestic political needs. In essence, the American policymakers were building their own international information gathering network that was to inform their own education policy needs. I also discussed the creation of a new assessment called the International Adult Literacy Survey that directly measured workers' functional literacy.

Domestically, American schools were governed and regulated through technocratic methods. The focus was on achieving objectives, improving
performance, rendering systems accountable to the public and measuring outcomes. In the American model of educational accountability, one turned to experts such as statisticians to provide policymakers with what was perceived as neutral and objective data. The predominant task became gathering information and data to inform education policy.

The OECD's claim as to the legitimacy of these international educational indicators relied on scientific expertise and technical knowledge. Public involvement was seen from the narrow vantage point of public relations and managing the media. The process of constructing educational indicators was depoliticized even though certain individuals were making moral, ethical and political decisions on what to measure and how to measure it.

A new power bloc formation was taking shape that governed and regulated educational systems through the compilation of international educational indicators. Educational indicators were used to measure and monitor educational systems. The revision of ISCED in 1997 created a more elaborate and less ambiguous classificatory system. In contrast to the indicators that were developed in the 1970s by the OECD, the new educational indicators that were developed in the 1980s had a new component: educational outcomes. These learning outcomes were represented by the results of student and adult performance on international tests.

It was an American model influenced by the ideas of the Republican party that came to be adopted internationally. American officials were able to exert their influence so that consensus was achieved for an evaluative and technical
understanding of education. Thus, the American model for problematizing education was institutionalized at the international level through the multilateral structure of the OECD. At the same time, I documented in this chapter how tensions between the OECD Secretariat and OECD member states over the control of the INES Project began to arise. These were rendered visible by the constant reorganization of the governance structure of the INES Project. In the next chapter, I examine the formation of the OECD Programme for International Student Assessment (PISA).
Introduction

In the previous chapter, I traced the creation of a new international student assessment that measured both science and mathematics achievement levels, the construction of international educational indicators at the OECD, and the creation of an international adult literacy assessment. I discussed how the international educational indicators project, the INES Project, was formed and the influential role played by the U.S. Department of Education in its formation. I also documented how the U.S. Department of Education approached the International Association for the Evaluation of Student Achievement (IEA) with a proposal for a new assessment in science and mathematics. The Third International Mathematics and Science Study (TIMSS) produced data for measuring comparatively the performance of American students and their progress towards becoming first in science and mathematics achievement by the year 2000.

In this chapter, I focus on developments within the INES Project that led to defining the need for a new OECD-based international student assessment. I map the creation of the OECD PISA in the 1990s.

Problematization of an OECD International Student Assessment

The INES Project was organized into several networks. One of the networks, Network A, was responsible for developing indicators on learning outcomes.
Network A members gathered data from various international assessments for the publication of *Education at a Glance* (EAG), an annual OECD report on educational indicators that was first published in 1992. Learning outcome data were compiled and derived from the IEA studies, the International Assessment of Educational Progress (IAEP), and the International Adult Literacy Survey (IALS).

In 1993, the Third International Mathematics and Science Study (TIMSS) came under criticism for being poorly managed. TIMSS was reorganized and a new director, Al Beaton, was assigned to lead the project. TIMSS continued to be criticized for the quality of its data by several Network A members. In March 1995, Network A invited Al Beaton and Bill Schmidt, director of the Survey of Mathematics and Science Opportunities that developed the TIMSS frameworks, to make presentations to Network A members on the status of the TIMSS study. Tjeerd Plomp, Chair of the IEA, was invited to discuss organizational changes taking place within the IEA. Scott Murray, International Coordinator for the International Adult Literacy Survey (IALS), was also invited to discuss the IALS assessment approach (OECD/INES 1995a).

At the March 1995 meeting which took place in Volterra, Italy, a series of decisions were made that eventually led to the segmentation of the IEA community of practice. The OECD Secretariat, which was represented by Tom Alexander, the Director of the Directorate for Education, Employment, Labour and Social Affairs (DEELSA), wanted full access to the IEA international database before it was made publicly available. DEELSA wanted to conduct its...
own data analysis in a timely manner. The IEA refused to comply with these demands. The IEA found itself compromising its autonomy as a non-governmental institution. DEELSA, in contrast, wanted access to regular and timely policy oriented data on educational outcomes. These two organizational entities could not arrive at a compromise. After the meeting, Tom Alexander decided that DEELSA would conduct its own assessment of student achievement to guarantee regular and timely data for its educational indicators (Interview Respondent #8, June 19, 2006; #9, 29 June 2006).

The Clinton Administration did not oppose this turn of events for several reasons. The U.S. Department of Education wanted an OECD-based study that would provide it with an additional yardstick for measuring student performance. The National Center for Education Statistics (NCES) would then be able to compare the performance of American students with the performance of students in OECD member states. U.S. Department of Education officials also believed that an OECD-based study would provide them with more direct control over the design of the assessment, which they did not have with the IEA studies. Lastly, with TIMSS facing organizational difficulties in the 1990s, it was better to have another reliable international data source that they could depend on (Interview Respondent #12, August 15, 2006; #27, August 27, 2006; #6, August 29, 2006).

The U.S. Department of Education also wanted to see a different assessment created that relied less on the school curriculum and more on the skills valued in the workplace. The Secretary of Labor’s Commission on
Achieving Necessary Skills (SCANS), which was established under the George H. W. Bush Administration in 1989, identified a set of competencies that students needed in order to be prepared for the world of work. These included basic skills in reading, writing and mathematics; thinking skills encompassing decision making, problem solving and creative thinking; and personal qualities involving individual responsibility, self-esteem and self-management. The Commission wanted to see American schools reformed so that they become high performing, thus “improving the match between what work requires and what students are taught.” With schools becoming high performance organizations, the Commission recommended that new administrative processes and assessments be put in place (SCANS 1992: 12).

The U.S. Department of Labor continued to implement these recommendations when the Clinton Administration took office in 1993. Secretary of Labor, Robert Reich, believed that in a globally integrated economy, “[e]ach nation’s primary assets will be its citizens’ skills and insights” (Reich 1992: 3). In fact, the Clinton Administration came into office with the intention of implementing a comprehensive education reform agenda (Riley 1995). Under Clinton, the National Education Goals of 1990 that were agreed upon by George H. W. Bush and the state governors were expanded and signed into law in March 1994 as Goals 2000: Educate America Act. Goals 2000 included a comprehensive reform agenda, the Safe Schools Act, and the reauthorization and reshaping of the Office of Educational Research and Improvement (OERI). The Clinton Administration established a cooperative
program between the Department of Education and the Department of Labor to improve the transition from school to work – the *School to Work Opportunities Act* became law on May 4, 1993 (Riley 1995: 7). The Clinton Administration was thus very interested in ensuring youth were equipped with the right skills for the world of work.

Given the Clinton Administration's policy agenda and its focus on workplace skills, international educational benchmarks were needed to measure progress towards Goals 2000. It was consistent with the Department of Education's overall objective to encourage two types of assessments – one geared to measuring student performance in relation to curricular objectives that was delivered by TIMSS and another aimed at measuring students' preparedness for the workplace that would be developed under the auspices of the OECD.

After the Volterra meeting of 1995, INES Network A members agreed that a new data source was needed for compiling regular indicators on student outcomes. The present international assessment studies (i.e. the IEA studies) were described as “too expensive, too slow to report, do not cover enough domains, do not involve enough OECD countries, and the OECD has little influence or control over the testing schedule” (Owen, Hodgkinson, and Tuijnman 1995: 216). Network A started working on a data collection strategy known as *A Strategy for Producing Student Achievement on a Regular Basis*. It was presented to the INES General Assembly in Lahti, Finland in June 1995. The Data Strategy elaborated an OECD approach to student testing that
combined curriculum based tests in reading, mathematics and science with cross-curricular competencies. The proposed assessment would target 15 year olds, since they would be at the end of their obligatory schooling and ready for the world of work. Work activities outlined in the Data Strategy's Implementation Plan were similar to other international student assessments, encompassing sampling procedures, pilot testing, data analysis and dissemination. The Data Strategy made several organizational recommendations. These included the creation of a decentralized governing body or steering committee with the OECD Secretariat responsible for administration and consensus-building; the contracting of evaluation work to an agency or to a consortium of agencies through a tendering process; and, the collection of national data to be conducted by participating states (OECD 1995). Overall direction of the Data Strategy would remain in the hands of the INES Project Network A.

The Data Strategy underwent several revisions as it was reviewed by more senior governing bodies within the OECD. Members of the Education Committee and the Centre for Educational Research and Innovation's (CERI) Governing Board wanted increased focus on measuring students' cross-curricular competencies. There had been a growing interest in the 1980s in measuring workers' functional skills. Both the U.S. Department of Education and the U.S. Department of Labor were especially interested in ensuring that students were graduating with the skills needed for the world of work. Network A had begun working on the feasibility of assessing students' cross-curricular competencies in 1991. These competencies were framed as skills young adults
needed to "play a constructive role as citizens in society." The areas of competencies included politics, economics and civics; problem-solving; self-perception and self-concept; and communication (OECD 1997a: 21). Further work was taken up by Swiss researchers in a project known as DeSeCo or Definition and Selection of Competencies.

It is interesting to note that there was surprisingly little dissent among members of Network A. American and European members agreed on the definitions of competencies that they were developing. As technical experts working for their respective ministries of education, their ultimate goal was to arrive at improving the technical aspects of measuring educational performance. Members of Network A worked closely with the IEA at developing these new definitions. Both American and European objectives overlapped as the Data Strategy evolved. European states had since the early 1990s, adopted a more instrumental approach to education. They were interested in improving the quality of their workforce and education was viewed as a tool for creating a competitive workforce. Even though the discourse from several European states emphasized social concerns, their primary objective remained "legitimizing and furthering programmes of change" (Jones and Duceux 2006: 97). An international educational assessment focusing on skills and competencies of students was an important tool for supporting new educational policy reforms that were aimed at developing a more competitive and highly skilled European workforce (Mitchell 2006).
In the final version of the Data Strategy, the problem domain was defined in terms of collecting "indicators on the knowledge, skills and competencies of students in reading, mathematics and science" (OECD/INES 1997d: 12, s.16). The U.S. Department of Education was happy with the problematization since it addressed its domestic requirements for comparing U.S. student performance with that of the performance of students from similar industrialized economies. European states were also happy with this definition since it served their needs for assessing the quality of their future workforce. The OECD was keen on moving forward with the Data Strategy as it supported its work on educational indicators and gave it renewed legitimacy in the area of education policy. Overall, OECD member states supported the project financially as it was a policy-oriented assessment that provided them with information on the performance of their education systems. Domestically, many of the member states had established accountability and evaluation frameworks for their schools so the data generated by the Data Strategy served as an additional information source to measure their progress towards high standards of student achievement.

With the Education Committee and the CERI Governing Board approving the final version of the Data Strategy, a Network A 'transition meeting' was called to order by Tom Alexander in May 1997 in order to transfer responsibility for the management of the Data Strategy to the standing committee that would oversee its implementation.

*First Meeting of the Board of Participating Countries: the Terms of Reference*
Network A members in an earlier meeting had decided to name the standing committee that oversaw the Data Strategy the ‘Board of Participating Countries’ (BPC). The BPC was a mechanism for ensuring compliance with policy priorities during the Data Strategy's implementation. Network A was to remain involved in the Data Strategy’s long term development (OECD/INES 1998).

The BPC held its first meeting in October 1997. Eugene Owen, the American delegate and Chair of Network A, was elected Chair of the BPC for a period of three years. An Executive Group was created with a Chair (Eugene Owen), three Vice-Chairs, and a representative from DEELSA. The Group's purpose was to make certain executive decisions between meetings of the BPC without consulting all the members of the BPC. The presence of Eugene Owen, who was a skillful negotiator, ensured American influence over the OECD-based assessment was maintained. A temporary name was given to the Data Strategy – the OECD Programme for Producing Student Achievement Indicators on a Regular Basis. It was not until July 1998, at the fourth meeting of the BPC, that the assessment came to be known as the OECD Programme for International Student Assessment (PISA) (OECD/DEELSA 1998b).

The Directorate for Education, Employment, Labour and Social Affairs (DEELSA) had already initiated work on the Terms of Reference for the Data Strategy. The Board of Participating Countries (BPC) reviewed the Terms of Reference which were based on Network A’s Data Strategy document and its Implementation Plan. The Terms of Reference outlined the project’s objectives and history. They detailed the project’s design, including all technical aspects,
such as content domains, survey cycle, target population, sample design, context questionnaires, and assessment instrument design. Included was a description of the management structure along with the roles and responsibilities of various participants, such as the BPC and the DEELSA.

The Terms of Reference stipulated a management structure consisting of functional expert groups in order to develop a "strong cognitive core and a coherent theoretical underpinning." The role of the functional expert groups was to "link the policy objectives specified by the BPC with substantive and technical expertise" and "to establish consensus on subject matter and technical issues among countries." The management structure also included the creation of a technical advisory group "to ensure the technical quality of the project" (OECD 1997b, s.60).

The Terms described the prime international contractor's role as having "ultimate decision-making authority and responsibility for operationalising the overall project design", for "implementing the project within the terms of reference and on time" and for "managing the process during implementation." The Terms ensured that the contractor had sufficient autonomy and flexibility in implementing day to day decisions pertaining to the project and in seeking expertise and contracting out work related to the project (OECD 1997b).

The Statement of Work that accompanied the terms of reference outlined the prime international contractor's tasks. It included a description of yearly activities at both the international and national levels for a period of four years. The tasks that it outlined included Framework Development, Instrument
Development, Sampling, Survey Procedures and Operations, Data Verification, Scaling and Preparing Data Product, Project Structure and Management. These work activities have long been established through the practices of the International Association for the Evaluation of Student Achievement (IEA).

The Framework Development described a process similar to that of the IEA studies. Test specifications were to be based on the curricula. Even though a cross-curricular component was included, the framework that the Terms of Reference and the Data Strategy broadly envisaged was very similar to that used in the IEA studies (OECD 1997b). Literacy, the conceptual framework that was to define the PISA, was to be formulated at a later time.

Because an OECD-based assessment was a governmental undertaking that was closely aligned with policy-oriented objectives, the test item development process was a highly politicized one. Each participating state wanted to ensure that it had test items that reflected its own curricula and educational goals. To appease the demands of participating states, the Terms of Reference identified a process involving item pool selection based on submissions made by participating states. By stating in advance the way in which item selection was to take place, the Terms of Reference avoided any discussion or deviation from this process. The politics involved in knowledge production and in building the infrastructure are manifested in the implementation of this procedural standard.

With the BPC approving the Terms of Reference, the tendering process was launched on October 10, 1997 and involved the submission of proposals for
implementing a student assessment. These proposals were also to include a breakdown of costs. As we shall see in the next section, the segmentation of the IEA community of practice takes place with the initiation of the OECD tendering process.

Segmentation of the IEA

Originally, there was one consortium formed that was to submit a proposal in response to the OECD call for tender. The IEA consortium was composed of several research institutes and centers led by Boston College, the operational centre of the IEA's TIMSS International Study Center. The members of the consortium had worked together on several IEA studies including the TIMSS. They did not view themselves as separate from the IEA since each institution contributed its expertise to the IEA studies. Work on each IEA study was delegated among research centers (Interview Respondent #14, 27 August 2006).

The Australian Council for Educational Research (ACER), one of the consortium members, was informally contacted by individuals from the Directorate for Education, Employment, Labour and Social Affairs (DEELSA) and from the Education Committee and told that the OECD found the American presence too dominant in the current consortium make up. ACER was informed that the OECD would like to see more of a European presence in the consortium and a different institution at its helm. ACER contacted the IEA consortium, shared this new information with the members, and suggested to them that an
ACER-led consortium might have a better chance of winning the tender. The OECD contacts had hinted to ACER that an IEA led consortium would likely be at a disadvantage (Interview Respondent #14, 27 August 2006; #19, 24 September 2006).

From the vantage point of the Australian Council for Educational Research (ACER), it was more pragmatic to take the lead, form a new consortium, and ask the IEA consortium members to join the ACER-led consortium. Boston College and Statistics Canada declined the offer to join the ACER-led consortium. ACER went ahead and formed its own consortium consisting of the Netherlands National Institute for Educational Measurement (CITO), the American-based Westat, and two experts from the University of Liège. The Japanese National Institute for Educational Research (NIER) and the Educational Testing Service (ETS) later joined the ACER-led consortium to reinforce politically ACER’s transnational representation. The IEA-Boston College-led consortium was now composed of the National Foundation for Educational Research for England and Wales (NFER) and Statistics Canada. (Interview Respondent #7, 9 June 2006; #14, 27 August 2006; #19, 24 September 2006).

The OECD’s political maneuverings encouraged the segmentation of the IEA community of practice, further deepening the fissures that were developing among its diverse membership. Curriculum experts, educational researchers and psychometricians were pursuing new areas of research work and measurement procedures. There was an interest in assessing student
knowledge in science based on a scientific literacy framework that incorporated an understanding of concepts, processes and values of science (Bybee 1997). New American mathematics standards were adopting a quantitative literacy approach to student knowledge (Schoenfeld 2002). As we saw in Chapter 6, adult functional literacy skills were being assessed using new frameworks. In Chapter 3, I discussed how, among psychometricians, a group of practitioners adopted item test measurement techniques that were developed by the Danish statistician, Georg Rasch. The Rasch model was popular among the psychometricians at ACER. The IEA was not able fully to accommodate these new problem domains, work practices and tools that practitioners involved in large scale assessments were eager to expand on. Furthermore, as more educational systems adopted standardized testing for addressing their accountability frameworks, demand for this type of expertise also grew, stimulating competition among research agencies.

In the next section, we shall see how an OECD-based student assessment facilitated an intersection of lines of work and tools, resulting in the formation of the PISA.

The Submission of Proposals and the Technical Review Panel

The OECD received three proposals in response to the call for tenders: one from the University of Bourgogne, a second from the ACER-led consortium and a third from the IEA-Boston College-led consortium.

A Technical Review Panel was created in order to review the proposals. The BPC nominated several experts to serve on this panel from which five were
selected. One of the panelists was Neville Postlethwaite who was a former Chair of the IEA. The panel met from November 20 to November 22, 1997 to review the proposals. They rated each proposal according to the following criteria:

- Organisational capacity and staff qualifications: 45 points
- Technical quality: 35 points
- Management capabilities: 10 points
- Past experience: 10 points

The Technical Review Panel presented its report at the second meeting of the BPC held in December 1997. The panel awarded the most points to the ACER-led consortium. According to the Panel’s report, the ACER proposal demonstrated “a clear understanding of the policy objectives” and addressed the tasks “thoroughly and in a creative manner” (OECD/DEELSA 1997c:1).

The expert panel was impressed with ACER’s approach to developing an integrated framework. The ACER proposal recommended creating functional expert groups that would simultaneously develop the frameworks in each domain. Such an integrated approach ensured that instruments would not need to be changed in the future and trends would be regularly monitored. The proposal critiqued the Third International Mathematics and Science Study (TIMSS) framework as too content-focused and recommended its review and extension by the functional expert groups. It pointed to the need for developing a different framework and test items at the outset, since the target populations of the TIMSS differed from those the OECD intended to target. If the OECD study had its own original framework, the ACER proposal noted, this would ensure the
right instruments were developed for the study without needing to modify them in the future. Such a process guaranteed the regular collection of trend data that the Data Strategy had emphasized (OECD/DEELSA 1997c).

In fact, ACER was distancing itself from TIMSS. It did not want to compete with TIMSS but preferred to develop a different approach that focused on measuring students' application of their learning. The ACER consortium wanted to generate new knowledge that was more oriented towards life skills and literacy rather than curriculum content (Interview Respondent #14, 27 August 2006). These areas of research reflected an alignment of research interests among members of the ACER-led consortium. An intersection was forming in which fields of research could be pursued within a broader evaluation initiative.

The expert panel found one of the strengths of the ACER proposal to be in its approach to building consensus among participating countries. The ACER-led consortium proposed to implement this consensus-building approach during framework and instrument development. Rather than ACER leading the intellectual undertaking, it would rely on the expertise of the Functional Expert Groups. These experts were not ACER staff members or ACER consortium partners but rather experts in the fields of reading, mathematics and science who would be contracted to do this work. The proposal noted that project decisions should be made by involving national representatives, outside experts and DEELSA. ACER's emphasis on coordination and consensus building
demonstrated its political astuteness and its understanding of the politicized nature of an OECD-based study (OECD/DEELSA 1997c).

The IEA-Boston College-led consortium recommended an approach similar to that of the TIMSS, and more generally, of the IEA studies. It was a more centralized decision-making process where final decisions with regards to the framework and the instruments were taken by Boston College and its partner, the British National Foundation for Educational Research (NFER). Consensus-building would occur late in the project rather than during its early phases. As the Technical Report phrased it, “essentially the BPC must rely on the resident expertise in this group” (OECD/DEELSA 1997c:14). The framework development process that the Boston College proposal recommended was to be pursued in a similar manner for the reading, science and mathematics domains.

The expert panel was impressed with the ACER-led consortium’s transnational capabilities. That this consortium involved educational research agencies in Europe, the Pacific and North America, was deemed to be one of its strengths. In contrast, the Boston College proposal was viewed as weak because it did not have a “non-anglophone partner” (OECD/DEELSA 1997c:17). In reality, individuals from the OECD Secretariat had recommended informally to ACER that its proposal would be better received if its consortium was made up of a diverse constituency (Interview Respondent #14, August 27, 2006).

The expert panel remarked on the ability of the ACER led consortium to “ensure wide international collaboration and representation” (OECD/DEELSA 1997c: 16). According to the expert panel, the Boston College proposal
"seemed stronger in building relationships between its constituencies" but, on
the other hand, the ACER proposal "seemed stronger in reaching out to
international expertise to be involved in the functional expert groups" (15).

Another strength the expert panel highlighted was the ACER proposal's
focus on quality control. It found its approach to be "more extensive ... by way
of site visits" and as demonstrating more "flexibility" and "transparency" than that
of the Boston College proposal. In contrast, the expert panel found that the
Boston College proposal did not sufficiently elaborate on quality control
procedures as compared with the ACER proposal (OECD/DEELSA 1997c).

In the area of technical quality, the two proposals recommended different
models for scaling the test items. Even though both applied the same
theoretical framework for building the test instruments, entitled Item Response
Theory (IRT), they varied in the way it was to be modeled. The Australian
Council for Educational Research's (ACER) proposal recommended the Rasch
model, since several psychometricians and statisticians at ACER were leaders
in this approach. The Boston College proposal preferred to apply a two or three
parameter model (OECD/DEELSA 1997c). As discussed in Chapter 3, two or
three parameter models allow for the inclusion of more difficult questions
(Interview Respondent #3, 24 November 2005). The experts on the technical
panel pointed out that both statistical approaches "lack[ed] a sound theoretical
foundation." They noted "establishing and describing proficiency levels should
be guided by an assessment framework" (OECD/DEELSA 1997c:19).
The Board of Participating Countries (BPC) made its decision following the Technical Expert Panel's presentation. The majority of the delegates on the BPC favored the proposal presented by the ACER-led consortium. Even though ACER had been a member of the IEA since the First Mathematics Study conducted in the 1960s, it was presented to the BPC as a separate entity (Interview Respondent #19, 24 September 2006). This probably helped sway the delegates towards choosing the ACER-led consortium since they wanted to undertake a more innovative and less complex design for comparing student achievement internationally (Interview Respondent #18, 27 September 2006).

The U.S. Department of Education would have preferred the Boston College-led consortium as the prime contractor for the OECD-based study. There was a strong collaborative relationship that was established with the IEA given that the TIMSS study center was based in Boston College (Massachusetts). Furthermore, the directors of TIMSS had worked in the past with the U.S. Department of Education. At the same time, the U.S. Department of Education had to agree with the Technical Review – the ACER proposal did address the requirements of an OECD international student assessment (Interview Respondent #6, August 29, 2006). The contract for the OECD-based student assessment, the PISA, was awarded to ACER.

When the Australian Council for Educational Research (ACER) and its consortium members officially became the prime contractor for the OECD, a division took place within the IEA community of practice. In its approach, ACER distanced itself from a content-based and curriculum oriented assessment and
developed a more literacy-based approach to assessing student knowledge. It also envisaged a different way for organizing work activities that emphasized consensus-building throughout the stages of the project, guaranteeing involvement of OECD member states in the decision-making process.

**Table 6: OECD PISA Developments**

<table>
<thead>
<tr>
<th>Date</th>
<th>PISA Developments</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 1995</td>
<td>After the meeting in Volterra, Italy, the OECD and Network A begin looking into creating an OECD based study of student achievement.</td>
</tr>
<tr>
<td>June 1995</td>
<td>Network A presents <em>A Strategy for Producing Student Achievement on a Regular Basis</em> at the INES Project General Assembly in Lahti, Finland.</td>
</tr>
<tr>
<td>May 1997</td>
<td>Network A 'transition meeting' to transfer responsibility for the management of the Data Strategy to the standing committee that would oversee its implementation – the Board of Participating Countries (BPC).</td>
</tr>
<tr>
<td>October 1997</td>
<td>BPC holds its first meeting. Tendering Process launched for project implementation.</td>
</tr>
<tr>
<td>December 1997</td>
<td>Technical Review Panel presents its final report to BPC. The BPC decides to contract ACER as its Prime Contractor for the PISA.</td>
</tr>
<tr>
<td>Nov. – July 1999</td>
<td>Field trials and pre-pilots take place.</td>
</tr>
<tr>
<td>Feb. – Oct. 2000</td>
<td>Administration of the assessment in participating states</td>
</tr>
<tr>
<td>December 2001</td>
<td>Results of first PISA cycle released – National and International Reports released to the public.</td>
</tr>
</tbody>
</table>

**Forming the OECD PISA**

As the International Association for the Evaluation of Student Achievement (IEA) community was segmented, an intersection took place that came to form a PISA community of practice. The intersection succeeded because the OECD's institutional structure was an effective interface for facilitating interaction and
communication. The OECD's institutional configuration supported the enrollment of community members into the new problem structure of international student assessment. Several of these members were American experts who had adopted a literacy approach to reading, mathematics and science education. They had built ties with their Australian and Dutch counterparts. American psychometricians had forged epistemic alliances when they embarked on the implementation of the International Adult Literacy Survey.

The intersection also succeeded because of the heterogeneous engineering that characterized the construction of the PISA. By heterogeneous engineering, I mean the constitutive process of inter-linking both human and non-human elements (Law 1987). Coordination among the diverse elements was done using a range of artefacts and tools. My research points to a persistent American influence in the process of heterogeneous engineering. American influence permeates the international educational measurement arena since American curriculum experts, educational researchers and psychometricians have a "headstart" in deploying these very same artefacts and tools in their national (NAEP) and state assessments (Djelic and Sahlin-Andersson 2006: 397). In the next section, I describe how intersection was effected.

**Institutional Structures**

The OECD's institutional structure facilitated the formation of an intersection of inter-related activities, tools and artefacts. Participants were being enrolled into this initiative as it was a new frontier in the assessment field. The OECD, a
governmental organization, could afford to commit itself to the new endeavor since it had the organizational, financial and political resources to ensure its success. Because of OECD institutional support, a new assessment was launched in 1997 that tested 15 year olds every three years. The frameworks were to be created and published in 1999. The first OECD PISA cycle was scheduled to take place in 2000, the second in 2003 and the third in 2006. Each cycle would test a major domain: reading was the major domain for 2000, mathematics in 2003 and science in 2006.

The most significant task the founders of an OECD-based assessment faced was the selection of subject matter and technical experts. The ACER-led consortium, now referred to as the International Consortium, the Board of Participating Countries (BPC) and the Directorate for Education, Employment, Labour and Social Affairs (DEELSA) began in December 1997 to pick the experts who would develop the assessment's domain frameworks and the technical experts who would oversee the sampling, statistical and psychometric components. The legitimacy of the OECD as an authority in the international assessment field rested on the expertise of these communities of practice which were enrolled into its network of associations.

There were three expert groups that had to be formed, one for each of the assessment domains: a Reading Expert Group, a Mathematics Expert Group and a Science Expert Group. The Chairs of these expert groups were chosen because of their domain expertise and their leadership in their respective fields. There was also a geo-political motivation for the choices
made so that representation included an American, a European and a British Chair. The Chair of the Reading Expert Group was Irwin Kirsch. Kirsch worked in the area of literacy assessment and had directed the International Adult Literacy Survey (IALS) for the Educational Testing Service. The Chair of the Mathematics Expert Group, Jan de Lange, was a leading Dutch subject matter expert in the area of mathematics and the Director of the Freudenthal Institute. He was chosen because of his work in Realistic Mathematics Education and authentic assessment that applied innovative techniques to learning and teaching mathematics (Interview Respondent #14, 27 August 2006). The Chair for the Science Expert Group, Wynne Harlen, was a science educator with extensive experience in researching and evaluating science learning.

Technical Experts were invited to become members of the Technical Advisory Group. Representatives on the Technical Advisory Group included members from the International Consortium and leading technical experts. These leading technical experts were also involved with the IEA studies. For example, Pierre Foy was an expert in statistical sampling; Larry Hedges, an expert in statistical methods for meta-analysis; and, Douglas Willms, an expert in the development of statistical methods for measuring Socio-Economic Status.

Many of the subject matter experts who gathered in 1998 to develop the domain frameworks had previously worked with one another. For example, several of the Reading group's members had worked together either on the International Adult Literacy Survey (IALS) or on various IEA studies. Those associated with the IALS were more inclined to create a framework that
assessed reading literacy skills needed for the world of work. Functional literacy test items were based on reading and interpreting newspapers, graphs, bus schedules, and labels. Two of the European experts believed the framework should reflect a subject-oriented approach to testing literacy that was academically grounded and literature based. The framework was to be the theoretical foundation out of which test items were to be developed. The actual production and selection of test items was a negotiated process that embodied a compromise among different approaches espoused by Reading Expert Group members (Interview Respondent #3, November 24, 2005).

The experts chosen for the Mathematics Expert Group and for the Science Expert Group also shared a desire to assess students in a non-traditional way that was not based on curricular knowledge. Most members on both expert groups had worked together on IEA-related studies, including the Third International Mathematics and Science Study (TIMSS). The American representatives on these expert groups, Thomas Romberg on the Mathematics Experts Group and Senta Raizen on the Science Expert Group, were actively involved in developing American national standards and assessments in these subject areas. Both Romberg and Jan de Lange, the Chair of the Mathematics Expert Group, worked together at the U.S. National Center for Improving Student Learning and Achievement in Mathematics and Science.

The experts who developed the frameworks were constrained by the requirements of the Board of Participating Countries (BPC), the policymakers who were overseeing the study's implementation. The experts felt that there
should be more open-ended responses. The BPC was concerned with the extra
costs incurred since open-ended responses required more intensive work for
correction, coding and statistical analysis. The BPC set the open-ended
responses up to 45% for reading and up to 35% for science and mathematics.
The rest of the item responses would be multiple choice. Another BPC
requirement was that all frameworks use the same language and concepts.
This helped create a more cohesive and integrated approach to the frameworks
(OECD/DEELSA 1998b).

Other requirements of the Board of Participating Countries (BPC)
reflected how political concerns influenced the construction of the PISA. A
recurring critique that members of the BPC voiced was cultural dominance or
cultural bias. They also complained about the biased representation of experts
within the frameworks. When the mathematics framework was reviewed by the
BPC, it was criticized for its Dutch bias and as “too narrowly oriented on Dutch
instructional models” (OECD/DEELSA 1998b: 5). The Dutch approach to
mathematics education, which espoused the philosophy of the mathematics
educator Hans Freudenthal, was not embraced by all members of the
mathematics community (Gravemeijer and Terwel 2000). Delegates on the BPC
were concerned that educational systems whose mathematics curricula were
built around Freudenthal’s philosophy of mathematics education might do better
on the mathematics test items than those whose curricula were not. The
mathematics framework was revised and resubmitted to the BPC for approval.
Another grievance voiced by BPC delegates was the 'Anglo-Saxon' bias that existed both in the language of the frameworks and among the representatives on each expert group. To address this criticism, the expert groups for the OECD PISA 2003 and 2006 were enlarged to include more expert representation. A process was put in place to allow for more expert involvement. The International Consortium created domain forums as a venue for expanding representation. For the PISA 2003, which had mathematics as a major domain, ACER organized a Mathematics Forum. Participating countries sent their own domain experts to participate in reviewing and finalizing the mathematics framework and the selection of test items. Similarly, a Science Forum was organized in 2005 in order to broaden the participation of the science literacy framework. These were political solutions that alleviated fears and criticisms of Anglo-Saxon dominance or cultural bias.

_Establishing Relations of Communication and Coordinating Work_

The intersection involved a process of heterogeneous engineering. Work practices, tools and artefacts were combined. Boundary objects and standardized packages created interfaces for intersecting lines of work. Standards, classificatory systems, and immutable mobiles were the glue, pulling the infrastructure together. In combination, infrastructural tools and work practices enabled action at a distance and control over peripheral communities such as national teams while simultaneously maintaining central control.

_Literacy – An Example of a Boundary Object_
At the third meeting of the Board of Participating Countries (BPC), ACER recommended the use of the term literacy for each of the domain names. The literacy concept helped make coherent and to integrate the framework definitions that were being developed. The concept acted as a boundary object that resolved coordination problems across groups doing work on the PISA. By boundary object, I mean an object that intersects several inter-related activities, helping to coordinate practices. The literacy concept that was introduced by the Australian Council for Educational Research (ACER) was “simultaneously concrete and abstract, specific and general, conventionalized and customized … and often internally heterogeneous” (Star and Griesemer 1989: 408).

Literacy defined internationally the outcomes that were being measured. It encompassed a broad interpretation of knowledge acquisition that included conceptual understanding of a subject domain, knowledge of processes, and application of this knowledge in various situational contexts. Such a broad interpretation of literacy in reading, mathematics and science also facilitated the language translation of the concept and its adoption at the national level. For example, in French, literacy was translated into the word ‘culture’ rather than the French word ‘alphabétisation’. Literacy was defined more concretely at the national level but these definitions varied according to cultural practices and conventions. In contrast, testing instruments require the use of standardized definitions so as to measure clearly defined abilities and test items.

Literacy as a boundary object was ineffective in changing local practices. Such a concept was too flexible and abstract to effect control over a localized
definition. In contrast, standardized methods delineating work activities and processes could change local practices.

**Standardized Packages**

The domain frameworks were essential for coordinating PISA work. They represented the conceptual maps for defining the knowledge being assessed, the test construction criteria and the interpretation of results for each of the domains. I use the concept of standardized package to describe the work that the frameworks did (Fujimura 1992). By standardized package, I mean an object that handles collective work across communities of practice while simultaneously doing standardization and fact stabilization work. The standardized package has a dual functionality. It contains boundary objects so as to allow for flexibility of application across groups and lines of work. It also has standardized methods leading to changes in local practices.

The frameworks as standardized packages contained boundary objects. Concepts such as ‘literacy’ were boundary objects that the frameworks defined. They also used abstract terms such as ‘content’, ‘competencies’, ‘skills’, ‘concepts’, ‘knowledge’, ‘situations’, and ‘tasks’ to categorize the various ‘dimensions’ of literacy. Boundary objects such as these could easily have diverse local interpretations. This may result in very different types of test items being written for the PISA instruments. I note this because the PISA instrument may be susceptible to not measuring what it is said to measure.

To avoid such ambiguity, the frameworks contained standardized methods and processes for item development. Test developers who were
submitting test items had a specified range of choices for item development. Each test item was to be written so as to assess a set of explicit tasks. For example, the ability to grasp scientific ideas was assessed based on the interpretation of a scientific text. The ability to retrieve information from a text or from a graph was assessed based on answers on a set of questions related to a newspaper article. Mathematical problem solving ability was assessed when a certain form of computation to solving the problem was applied (OECD 2000).

Standardized methods for writing PISA test items differed from methods applied for writing curricular or subject-based items. Test developers, who for years had written items for national assessments, were unfamiliar with these new methods. In order to write PISA destined test items, test developers needed to adopt new practices. Test developers were forced to incorporate new ways of writing test items that did not rely on curriculum content. Test developers were writing items that assessed specific competencies as defined by the domain frameworks rather than content-based knowledge (Interview Respondent #20, October 20, 2006; #21, October 25, 2006).

As standardized packages, the frameworks served several purposes for various community members and users. The domain experts and the International Consortium used the frameworks to assist them in writing the test items, in reviewing items submitted by national teams, and in rationalizing reasons behind selection or exclusion of items. The frameworks were used as guidelines to write test items by national teams. The OECD used the frameworks to differentiate its assessment from previous international studies.
conducted by the IEA. The Board of Participating Countries (BPC) used the frameworks to achieve policy-oriented objectives that were closely linked with the improvement of their 'human capital stock' (OECD 1999).

Standards, classificatory systems, and immutable mobiles

The International Project Centre was the technical centre for the OECD PISA. All National Centres or the research centres representing participating countries at the national level, reported to the International Project Centre. A National Project Manager acted as the link between the International Project Centre and the participating country. National Centres implemented other international assessments such as those relating to adult literacy (IALS) and IEA studies. National Project Managers mediated between the local school level, the national level and international level of the project. At the local level, School Coordinators were assigned to each school and Test Administrators were trained to administer the PISA assessment. These National Project Managers came into existence with the IEA studies.

There was a multitude of activities involved at the national level in the implementation of all international student assessments, including the OECD PISA. Test Administrators were hired to administer the tests. School Coordinators were employed to coordinate between the school and the National Centre. For students who were non-English and non-French speakers, translators were hired to translate the PISA instruments (test booklets and the questionnaires). There were 9 test booklets for PISA 2000 and 13 test booklets for PISA 2003. Test booklets and questionnaires had to be assembled,
organized into packages and sent to the respective school coordinators. Markers were hired and trained to mark the tests.

Standards hold the OECD PISA together by ensuring all work is conducted in the same manner across sites. These include procedural instructions and manuals for test administration and school coordination; test schedules that are updated and disseminated from one centralized location; training manuals for test administrators and school coordinators; and procedures for organizing test materials for coding and data entry. The PISA is deemed to be of high quality precisely because it has so many quality standards in place.

The PISA deployed two classification systems to render the data comparable. As boundary objects, classification systems helped coordinate work, since they embodied abstract concepts allowing for concrete implementation through categorical work. For example, educational programmes in which the students were enrolled were classified using the International Standard Classification of Education (ISCED). The students were asked to identify their parents' levels of education and occupation. These were then reclassified using ISCED for levels of education and the International Standard Classification of Occupations (ISCO) for occupations. Teachers' levels of education were also coded using ISCED. The information was derived from questionnaires that the school principals filled out. The classification of the data required categorical work to be done. The appropriate categories have to be first identified by those individuals filling out the questionnaires; then, categorical work has to be done so that the information on the questionnaires is fitted into
the classification scheme. Despite attempts at universalizing the reporting of phenomena, there are always contingencies that arise whereby information is locally interpreted (Bowker and Star 1999).

Immutable mobiles are ubiquitous and essential for maintaining continuity within the PISA infrastructure. By immutable mobiles I mean two dimensional inscriptions or representations that carry information. Immutable mobiles collect information from the peripheries, the National Centres, and bring the information back to a centralized location, the International Project Centre. They can then be recombined to produce new representations. Immutable mobiles link together heterogeneous practices of the PISA infrastructure and build associations between human and non-human entities. Each immutable mobile serves a particular purpose and represents a set of practices.

For example, inscription devices such as the List of Schools and the Student Tracking Form were used for administrative and metrological purposes. The List of Schools contained the PISA population size (i.e. enrolled students), the grade population size and the sample size. In statistical terms, it was referred to as the sampling frame. The student sample was selected from the List of Schools. The sample student population from each school was transcribed onto the Student Tracking Form, described as the “central administrative document for the study.” This standardized form was sent to each participating school so that the school had a record of the complete list of its student sample. The Student Tracking Form was used to link both the achievement tests (called ‘booklets’) and the Student Questionnaires to the
students when they received all the assessment's instruments. It also was used to input 'exclusions' for those students excluded from participation in the assessment. These were coded as 'functionally disabled', 'educable mentally retarded', 'limited test language proficiency', and 'other'. Students who did not participate were coded as: 'absent', 'present for part of the session', 'parent refusal' and 'other'. The Student Tracking Form was important in calculating total exclusion rates for each national student population (Adams and Wu 2002; OECD/PISA 2001).

The Student Tracking Form was used to check that all the tests and questionnaires were collected for each student and for each school. The Student Tracking Form, the tests, and the questionnaires were sent to the National Centre. Markers marked the tests and filled out marker record sheets. Questionnaire responses were entered into a computer database. All these inscriptions were then transferred onto data files and submitted electronically to the International Project Centre for processing. The International Project Centre then cleaned the data, plugged the data into several statistical models for weighting, scaling and for statistical analyses. The data then were sent back to the National Centres so that they could produce their own national reports.

The International Consortium proceeded to write an international report based on the data it had collected. Once the OECD PISA results were released, they then became available for downloading from a centralized on-line database. The PISA data were recombined into reports and publications destined to be consumed by policy makers, educators, researchers, the media
and other individuals involved in the education process. Immutable mobiles in the form of OECD publications were essential for information dissemination.

Through the deployment and construction of tools, artefacts, standards and classification systems, the community of practice of the PISA created an infrastructure for a new student assessment. Relations of communication were established among community members and coordination of work using various tools and inscription devices reconciled divergent viewpoints and approaches.

**Figure 5: An example of an immutable mobile - the Student Tracking Form**

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Grade</th>
<th>Birthday</th>
<th>Excluded</th>
<th>Code</th>
<th>Booklet</th>
<th>Number</th>
<th>Original Session</th>
<th>Attendance Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EXCLUSION CODES** (Col. 7): **PARTICIPATION CODES** (Cols. 9A, 10B)

1 = Functionally disabled  0 = Absent
2 = Educable mentally retarded  1 = Present
3 = Limited fine language proficiency 8 = Not applicable, i.e., excluded, no longer in school, not age-eligible
4 = Other

**Conclusion**

The chapter provided a detailed discussion of the formation of the OECD PISA. I began with a discussion of how a new problematization was initiated and a Data
Strategy was developed for a new international student assessment. I revealed the negotiations that took place to arrive at a final conceptualization of the kind of international assessment that was desirable for OECD member states. I discussed the tendering process that was initiated by the OECD and the segmentation of the IEA. I elaborated on the technical review of the proposals that were submitted to the Board of Participating Countries and explained how a literacy approach came to be introduced into the problematization of the PISA. I concluded that the cohesiveness of the intersection that took place is attributable to the strong institutional structure of the OECD as a governmental organization. At the same time, the formation of this power bloc is a reflection of an adjustment that has taken place among technical capacity, relations of communication and power relations. If one of these relationships begins to falter, then this system of regulation could lose its ability to exercise power.

In the next chapter, I discuss the power relations of both the PISA and the IEA studies.
Chapter 8
Power Relations

Introduction
In earlier chapters, I have focused my analysis on the PISA power bloc's technical capacity and relations of communication. In the previous chapter, I explained how the International Association for the Evaluation of Student Achievement (IEA) community was segmented and out of this, a new PISA community of practice was formed. I looked at the coordination work of boundary objects, standardized packages, standards, classification systems and immutable mobiles that helped coordinate work and build relations of communication for the PISA.

In this chapter, I focus on the relationships of power that constitute the bloc formation. There are two aspects to power relations I discuss. The first involves the power to govern the conduct of individuals through the categorisation and measurement of student learning. I examine the codification of the object 'literacy.' I analyze the discourse of official knowledge and its use of 'literacy' to legitimize the present economic order. I point out the myths that the official knowledge perpetuates. The second aspect of my analysis of power relations focuses on institutional structures. I compare the OECD and the IEA as institutions which attempt to exert their power to control knowledge production and knowledge exclusivity in the realm of international educational indicators.
Governing Conduct

We have examined how literacy as a boundary object facilitates the coordination of work across communities of practice. I would like now to examine how literacy as an object is used by those in power to govern the conduct of individuals. I suggest that 'literacy,' as it has been conceived by the OECD PISA, is as an object that is very similar to the object 'intelligence.' It is a psychometric construct that serves to objectify and quantify the skills and competencies of the knowledge worker in the global economy. The quantification of literacy does violence to both teaching and learning as it facilitates the instrumental use of the school for the purposes of the neoliberal project and standardizes learning, a creative and experimental process, into a series of decontextualized and individualized tasks.

In order to understand how international student assessments attempt to govern the learning process, it is useful to compare the IEA studies to the OECD PISA and observe how they construct their respective objects for measuring student knowledge. The IEA conducts two sets of assessments that can be compared with the PISA. The first is the Progress in International Reading Literacy Study (PIRLS)\(^{79}\) and the second is the Trends in International Mathematics and Science Study (TIMSS). The IEA approach has been to use participating states' curricula as a framework for developing IEA test items. I suggest that the testing approach of the IEA studies reinforces the governing of

---

\(^{79}\) In 1970, IEA conducted its first reading comprehension study as part of the Six Subject Study. In 1991, it implemented the IEA Reading Literacy Study.
students, teachers and schools by the state since it is closely associated with national curricula. In contrast, the OECD PISA does not follow a particular curriculum but has constructed the literacy object consisting of three areas: reading literacy, mathematics literacy and science literacy. One of the interview respondents explained, “They [OECD PISA] are advocating a different outcome of the curriculum – but they are not saying here is what you have to teach” (Interview Respondent #21, 25 October 2006). Yet by determining these curricular outcomes, the OECD PISA is identifying what students need to learn in order to be prepared to work in the global economy.

*Constructing the Instrument to Measure the Literacy Object*

In order to understand how the literacy object was constructed, I will first return to the International Adult Literacy Survey (IALS). The IALS assessed adults aged sixteen to sixty-five. The IALS report written in 1996 noted that “literacy is now equated with an individual’s ability to use written information to function in society” (11). The IALS codified this vague concept of a ‘people’s ability to function in society’ by assessing their abilities in three areas or domains: prose, document and quantitative (see Table 7). Each area or domain had a corresponding scale of zero to five hundred. The scale encompassed various levels of difficulty ranging from Level 1 to Level 5 (see Table 8).
Table 7: The Literacy Domains in the IALS

<table>
<thead>
<tr>
<th>Domain</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prose literacy</td>
<td>the knowledge and skills needed to understand and use information from texts including editorials, news stories, poems, and fiction</td>
</tr>
<tr>
<td>Document literacy</td>
<td>the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables, and graphics</td>
</tr>
<tr>
<td>Quantitative literacy</td>
<td>the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a checkbook, calculating a tip, completing an order form, or determining the amount of interest on a loan from an advertisement</td>
</tr>
</tbody>
</table>

Table 8: Literacy Scale for Prose, Document and Quantitative

<table>
<thead>
<tr>
<th>Level</th>
<th>Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 to 225</td>
</tr>
<tr>
<td>2</td>
<td>226 to 275</td>
</tr>
<tr>
<td>3</td>
<td>276 to 325</td>
</tr>
<tr>
<td>4</td>
<td>326 to 375</td>
</tr>
<tr>
<td>5</td>
<td>376 to 500</td>
</tr>
</tbody>
</table>

Table 9: IALS Literacy Abilities

<table>
<thead>
<tr>
<th>Proficiency Level</th>
<th>Description of abilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Persons with very poor skills. For example, in a prose literacy task, respondents are asked to determine the amount of medicine to give a child from information printed on package.</td>
</tr>
<tr>
<td>Level 2</td>
<td>Persons are able to deal with material that is simple, clearly laid out, and in which tasks involved are not too complex. Difficulties in facing novel demands such as learning new job skills. For example, in a prose literacy task, respondents are asked to read a page from a bicycle owner’s manual and summarize how to ensure the bicycle seat is in the correct position.</td>
</tr>
<tr>
<td>Level 3</td>
<td>Suitable minimum for coping with demands of every day life, working in complex, advanced society. Requires ability to integrate several sources of information and solve more complex problems. For example, in a document literacy task, respondents are asked to compare two graphs on sales and injuries and analyze their contents.</td>
</tr>
<tr>
<td>Levels 4 and 5</td>
<td>Persons who demonstrate higher-order information processing skills. For example, in a quantitative literacy task, respondents are asked to look at a nutritional analysis table and determine the percentage of fat that comes from a Big Mac.</td>
</tr>
</tbody>
</table>

The test subjects were required to complete the test instrument which was composed of several tasks varying in difficulty (see Table 9). Based on this evidence, the data was analyzed to determine the adult population's literacy abilities. The tasks the test respondents completed were representative of how dominant discursive practices and a consumer oriented economy define literacy. The tasks are decontextualized from the "multiple and shifting purposes of individual citizens and their communities" (Hamilton 2001: 179).

The IALS reports analyzed the results as evidence of the association of literacy skills with workers' employment prospects. Nevertheless, studies have pointed to how employers value certain behavioural and personality traits over cognitive traits. There were other variables at play, other than literacy skills as constructed by the IALS, which were attributable to reaping the rewards of the labour market (see Bowles and Gintis 2002: 10-12).

The creators of the OECD PISA looked to the IALS instrument as their model. It provided a template that they could build on to help them create an instrument for measuring the knowledge and skills of students who they defined as ready for the world of work.

The literacy object that the creators of the PISA wanted to measure was equated in the OECD discourse with human capital formation. In a 1998 report published by the OECD's Centre for Educational Research and Innovation (CERI), human capital was defined as "the knowledge, skills, competences and other attributes embodied in individuals that are relevant to economic activity" (OECD/CERI 1998: 9). The CERI report argued that there was "a large gap
between existing evidence and direct measurement of human capital.” It went on to say that this gap was “being closed through major measurement exercises being launched by the OECD.” At the time, these measurement exercises represented surveys being developed to quantify students’ “cross-cultural competencies” and adults’ “life skills” (11).

The CERI report defined in economic terms what a measure of human capital was required to do.

To produce more direct measures of human capital, information about both individuals and settings need to be collected. For individuals, clarification is needed empirically of the knowledge, skills, competences and other attributes that enhance productivity, innovation, and employability in different employment-related situations. These imply a large international research agenda. Measures are needed not only of stocks at any one time, but of rates of appreciation and depreciation over time. Such measures should be designed to identify the organisational and economic conditions in which human capital is most likely to be built on or lost (11).

In order to ensure that school systems are producing the right type of worker, not only do the conduct of learners and teachers have to be governed but also schools have to be governed to ensure that students are taught the appropriate skills and competences for the world of work.

As we saw in the previous chapter, using literacy as a measure of human capital was first suggested by ACER as a means for measuring student achievement. The PISA’s first publication which introduced the assessment framework explained that “the indicators are designed to contribute to an understanding of the extent to which education systems in participating countries are preparing their students to become lifelong learners and to play...
constructive roles as citizens in society" (OECD 1999a). The discourse was more concerned with governing the conduct of students as lifelong learners and as citizens and did not reflect CERI’s economic tone. Yet the underlying objective that unified OECD member states was to be able to measure the international competitiveness potential of their labour force in a knowledge based economy. The assessment framework quantified the literacy of the knowledge worker in terms of reading literacy, mathematics literacy and scientific literacy.

Quantifying Reading Literacy

As the PISA framework report explains, the PISA’s reading literacy domain builds on and extends the definitions of literacy used by both the IEA Reading Literacy Study80 of 1991 and the IALS.81 “Reading literacy is understanding, using, and reflecting on written texts, in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society” (OECD 1999a: 20). According to the OECD PISA, its definition emphasises the notion that reading is interactive and requires reflection. It also includes a range of situations in which literacy takes place. So, how does the PISA reading literacy instrument manage to codify and quantity this definition? The creators of the PISA constructed three sub-scales for reading literacy, each with its own set of task variables. The PISA 2000 Technical Report notes that “[t]hese variables are based on the

80 The IEA Reading Literacy Study defined literacy as: “the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers, and for enjoyment (quoted in Kapinus 2003: 3).

81 The IALS defined literacy as: “Using printed and written information to function in society, to achieve one’s goals, and to develop one’s knowledge and potential” (OECD 1999a: 19)
judgements by reading experts of the items that fall within each level band rather than on any systematic analysis" (Adams and Wu 2002: 200, my emphasis).

The three reading sub-scales were ‘retrieving information,’ ‘interpreting texts,’ and ‘reflection and evaluation.’ I will discuss in greater detail the third sub-scale, ‘reflection and evaluation,’ since the Technical Report identified it as “PISA’s major new contribution to conceptualising reading.” The Technical Report notes, “PISA is the first large-scale study to attempt to define variables operating within the reflection and evaluation sub-scale and to propose characterising populations from this perspective in its reporting” (200). The reflection and evaluation sub-scale categorizes this process into five levels of difficulty.

Table 10: Measuring PISA Reading Abilities – Reflection and Evaluation Sub-Scale

<table>
<thead>
<tr>
<th>Level of difficulty</th>
<th>PISA Reading Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Make a simple connection between information in the text and common, everyday knowledge. The reader is explicitly directed to consider relevant factors in the task and in the text.</td>
</tr>
<tr>
<td>Level 2</td>
<td>Make connections or comparisons between the text and outside knowledge. Draw on personal experience and attitudes to explain a feature of the text. The tasks require a broad understanding of the text.</td>
</tr>
<tr>
<td>Level 3</td>
<td>Make connections, comparisons and explanations, or evaluate a feature of the text. Demonstrate a detailed understanding of the text in relation to familiar, everyday knowledge. Draw on less common knowledge. The reader may need to infer the factors to be considered.</td>
</tr>
<tr>
<td>Level 4</td>
<td>Critically evaluate a text, or hypothesise about information in the text, using formal or public knowledge</td>
</tr>
<tr>
<td>Level 5</td>
<td>Critical evaluation or hypothesis, may draw on specialised knowledge; deep understanding of low and complex texts. Example: The student takes account of inconspicuous as well as more obvious information in a complex text on a relatively unfamiliar topic</td>
</tr>
</tbody>
</table>

Source: Adams and Wu 2002: 207.
In comparison, PIRLS 2001 assessed 4th grade students in two areas of reading: literary experience and acquiring and using information. However, the process differed from the PISA’s approach to assessing reading literacy because national curricula and national teaching approaches were considered.

Reading comprehension was organized into four areas of ability:

- focus on and retrieve explicitly stated information
- make straightforward inferences
- interpret and integrate ideas and information
- examine and evaluate content, language, and textual elements (Mullis et al. 2003: 282)

Furthermore, the IEA compiled an encyclopaedia of each state’s educational system with a focus on its approach to reading. The *PIRLS 2001 Encyclopedia* was published as an additional resource to the *PIRLS 2001 International Report*.

That is not to say that the IEA is better or worse than the OECD PISA. I wanted to point to how the IEA supports state-led educational objectives that govern the conduct of citizens within national boundaries. Both the IEA and the OECD PISA approaches to assessing students internationally are integrated into the broader network of global educational relations. The data they generate are beneficial to those agents within the global network who monitor and survey the performance of educational systems.

*Quantifying Mathematics Literacy*

As noted in the previous chapter, the mathematics literacy framework was influenced by the Realistic Mathematics Education approach developed by Hans
Freudenthal, the same individual who criticized the IEA studies back in the 1970s. PISA defined mathematics literacy in the following way:

Mathematical literacy is an individual’s capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgements and to engage in mathematics, in ways to meet the needs of that individual’s current and future life as a constructive, concerned and reflective citizen (OECD 1999a: 41).

The framework emphasized that the PISA approach to measuring mathematics literacy was founded on “assessing the full breadth of student achievement in a coherent, integrated way, rather than to test fragmented pieces of factual knowledge.” It highlighted the fact that it was different from the IEA TIMSS, an assessment based on “the curricula of participating countries.” The authors of the mathematics framework noted that, “[a]s result, the IEA/TIMSS instruments pertained mostly to knowledge of mathematical facts that were tested in isolation, mainly with very short items” (OECD 1999a: 48). So, how does the PISA codify mathematics literacy? The PISA mathematics literacy scale was organized into three broad level – from the lowest mathematics abilities to the highest mathematics abilities (see Table 11).

Quantifying Scientific Literacy

The PISA framework defined scientific literacy as: “The capacity to use scientific knowledge, to identify questions and to draw evidence based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity” (OECD 1999a: 60). Codification of this literacy definition into the scale was implemented from the lowest scientific abilities to the highest scientific abilities (see Table 12).
### Table 11: Measuring PISA Mathematics Abilities

<table>
<thead>
<tr>
<th>Proficiency Level</th>
<th>Examples of PISA Mathematics Tasks</th>
</tr>
</thead>
</table>
| Lowest Level      | • Recognise familiar elements in a problem, and recall knowledge relevant to the problem  
                   • Reproduce known facts or procedures to solve a problem  
                   • Solve problems involving only one or two steps |
| Middle Level      | • Interpret, link and integrate different information in order to solve a problem  
                   • Use and manipulate given mathematical models to solve a problem and use symbolic language to solve problems  
                   • Solve problems involving a small number of steps |
| Highest Level     | • Show insight in the solution of problems  
                   • Develop a mathematical interpretation and formulation of problems set in a real-world context  
                   • Solve problems involving several steps  
                   • Reflect on results and generalise findings  
                   • Use reasoning and mathematical argument to explain solutions and communicate outcomes |

Source: Adams and Wu 2002: 209

### Table 12: Measuring PISA Scientific Abilities

<table>
<thead>
<tr>
<th>Proficiency Level</th>
<th>Examples of PISA Scientific Tasks</th>
</tr>
</thead>
</table>
| Lowest Level      | • Recall simple scientific factual knowledge (e.g., names, facts, terminology, simple rules)  
                   • Use common knowledge in drawing or evaluating conclusions. |
| Middle Level      | • Use scientific concepts in making predictions or giving explanations  
                   • Recognise questions that can be answered by scientific investigation and/or identify details of what is involved in a scientific investigation  
                   • Select relevant information from competing data or chains of reasoning in drawing or evaluating conclusions |
| Highest Level     | • Create or use simple conceptual models to make predictions or give explanations  
                   • Analyse scientific investigations in relation to, for example, experimental design, identification of idea being tested  
                   • Relate data as evidence to evaluate alternative viewpoints or different perspectives  
                   • Communicate scientific arguments and/or descriptions in detail and with precision |

Adams and Wu 2002: 212

In comparison to the PISA, the IEA developed the TIMSS using a curriculum framework. Students’ achievement in mathematics and science was
assessed by developing test items that reflected the content of the national educational curricula. TIMSS results reflected performance levels in relation to these content areas. TIMSS 1999 results were scaled using a new approach— that of 'international benchmarks.' TIMSS researchers identified four points on the scale that were used as international benchmarks so as to compare performance among various student populations. The benchmarks were set at 10%, 25% (Upper quarter benchmark), 50% (Median benchmark) and 75% (Lower Quarter Benchmark). For example, international benchmarks would be applied in the following manner with respect to student performance in mathematics:

If student achievement in mathematics were distributed in the same way in every country, then each country would be expected to have approximately 10 percent of its students reaching the Top 10% Benchmark, 25 percent the Upper Quarter Benchmark, 50 percent the Median Benchmark, and 75 percent the Lower Quarter Benchmark (Mullis et al. 2000: 39).

Benchmarking is a technique developed in the private sector and applied as part of the New Public Management practices. According to the Oxford English Dictionary, a benchmark is used to evaluate or check performance against a standard (2007). In this case, student performance on the curriculum is set to an international standard that is based on the combined results of states participating in 1999 TIMSS (Mullis et al. 2000).

What I find revealing in this exercise of identifying how the PISA codifies its literacy definitions is that a curriculum outline emerges for each of the three domains. The PISA’s standardized tests of achievement become the templates...
for such a curriculum. The objective of a standardized, global curriculum is to produce the knowledge worker who is destined to work in the global economy. The variables chosen to assess the skills and competencies of this worker are established based on the values and judgements of a select group of experts.

If one took the various tasks for each domain and developed a curriculum according to those abilities, then one can fairly assume that students who are taught such a curriculum would do well on the PISA. In contrast, those who are taught a curriculum that does not reflect these abilities would not do well on the PISA.

In effect, what we have is an international organization that has defined the role of schools as producing the skilled worker for the knowledge-based economy. It has set the standard for measuring the performance of schools by creating an international student assessment that is based on an ambiguous object called ‘literacy.’ Whereas the IEA measured student performance according to national curricula, the OECD has created an assessment that measures student competencies for the global economy. It has set a standard for a global curriculum. Instead of schools producing workers for the local labour market, schools are now conceived as producing workers for the international labour market whose skills are to be assessed at age 15. In defining the student population to be 15 years of age, the experts and policymakers associated with the OECD PISA have determined this to be the “age at which educational destinies are settled” (Jones and Duceux 2006: 101).
Official knowledge has been legitimized through the process of constructing and codifying an object called literacy. The construction of a new international assessment has implications for how states choose to teach, evaluate their students, and design their curricula. Michael Apple points out the following:

What counts as knowledge, the ways in which it is organized, who is empowered to teach it, what counts as appropriate display of having learned it, and – just as critically – who is allowed to ask and answer all these questions, are part and parcel of how dominance and subordination are reproduced and altered in this society (1996: 22).

The official knowledge that the OECD produces tells us that we need more knowledge workers in the knowledge based society. In fact, what is happening in OECD member states is that as welfare programs have been cut, individuals have turned to the labour market to seek living wages. The problem is that “[t]here are simply not enough good jobs to go around” for everyone. Therefore, the stakes for achieving labour market success have become higher. In certain economies, job skill levels have not even improved: “The American and UK economies have failed to raise skill levels throughout the economy. They are characterized by enclaves of ‘knowledge’ work alongside swathes of low-wage, low-skilled jobs” (Brown 2003: 150).

One wonders why there is a need for schools to produce so many knowledge workers. Is it blind faith in the market’s ability somehow to create new knowledge jobs in the future? Do the jobs that are being created reflect increased demand for skills and greater on the job creativity or greater control and demand for efficiency? Or, does the production of more knowledge workers...
supply a highly skilled reserve army of labour that will be accessible to transnational capital at very low wages (Huws 2006). Perhaps PISA results simply serve as a lever, as ammunition, to jolt states into a path of educational reform, particularly when teachers' unions and the public are against these reforms. PISA numbers can be held up to legitimize the “dismantling of the curricula and pedagogies which sustained the uneven advances of the post-war decades, in favour of a model whose most notable features are sharp social segregation, strong managerial direction and an emaciation of teaching and learning” (Jones and Duceux 2006: 106).

I now move to the second aspect of power relations discussed in this chapter. The discussion pertains to the institutional structures and relations that circulate official knowledge.

**OECD Institutional Structures**

Power relations inside an international organization such as the OECD can be analyzed by examining the roles that the OECD performs. The OECD functions as policy instrument, as a policy forum and as a policy actor (Lingard 2000). As a policy instrument, the OECD becomes a tool for national policymakers and politicians to bolster certain policy directions and to legitimate policy decisions. The OECD as a policy forum becomes an arena where delegates come together, discuss policies, share best practices and in certain cases set future policy agendas behind closed doors. As an organizational actor, the OECD is a player in the international arena with its own organizational mandate and
objectives (Lingard 2000; Martens and Balzer 2004; Salzman and Terracino 2006). As policy actor, the OECD is concerned with distinguishing itself from other organizational actors and consultancies and in maintaining its relevance (Marcussen 2004). Tensions arise as the OECD attempts to reconcile these competing roles and functions.

The Education Directorate

In 2002, the OECD created a separate Directorate for Education "to enhance the OECD's work in this area and to better respond to the needs of citizens and governments" (McGaw 2002: 3). This organizational realignment elevated the status of education within the Education Directorate and sent a message to all the OECD member states that education policy was important to the success of economic policy. The creation of a separate Education Directorate also attested to the administrative complexity required for implementing and coordinating various education-related initiatives such as the Indicators of Education Systems (INES) Project and the PISA. With an annual expenditure of EUR 14 million (approximately US$ 18 million), the Directorate of Education is by no means a small player within the OECD (OECD/EDU 2005a: 5).

Within the OECD's organizational structure, it is the Education Directorate that is responsible for all OECD education related activities and programs. The Education Directorate's role as a legitimate conveyer of educational information, statistics, and policy advice is strengthened when it can exert control over the knowledge created and produced by the PISA. The Education Directorate's Indicators and Analysis Division, which is headed by Andreas Schleicher, is
responsible for managing and administering the PISA and the INES Project. The Education Directorate manages the knowledge created and produced by the PISA.

There is a tension within the OECD as it addresses its needs as policy actor and its member states' needs as a policy forum and policy instrument (Reinalda and Verbeek 2004). The Education Directorate would like to exert control over ideas and practices flowing through the organization so it can improve its knowledge creation and production capacity. I suggest that it deploys reorganization and restructuring mechanisms as tools to effect control. I provide several examples that demonstrate how such control takes place. At the same time, the Education Directorate serves the needs of member states through its committee work. The OECD, as an international organization, also has to fulfill its role as a policy forum and policy instrument.

The Education Committee, the policy forum for representatives from ministries of education of OECD member states, governs three of the seven education-related initiatives within the Directorate for Education. These include the work undertaken by the Directorate's Education and Training Policy Division and the Unit for Co-operation with Non-Member Economies. These programs have Part 1 funding, which means that they are funded by the OECD's central budget. The Education Committee also governs the INES Project jointly with the Centre for Educational Research and Innovation (CERI) Governing Board. The INES Project receives both Part 1 funding and Part 2 funding. The Part 2
funding means that the project receives financial contributions from those states participating in it (Interview Respondent #19, 24 September 2006).

**Figure 6: Education at the OECD - Committee & Governing Board Structure**

The INES Project was a forum for technical experts rather than a policy forum. Its objective was to produce indicators for policymakers to help inform policy decisions. The INES Project reflected the research agendas of OECD member states in their respective ministries of education. For example, at Network A meetings, members compared the way in which their ministries created and administered student assessments. It was a forum for exchanging technical knowledge with the majority of the funding coming from the U.S. Department of Education.

The other four educational initiatives are overseen by their respective boards. These include the PISA, the Programme on Educational Building
(PEB), the Programme on Institutional Management in Higher Education (IMHE), and the Centre for Educational Research and Innovation (CERI). These initiatives receive Part 2 funding since they are not OECD initiated programs, but were created as special projects.

In 2004, the Board of Participating Countries was renamed the PISA Governing Board. At the 17th meeting of the Board of Participating Countries (BPC), Andreas Schleicher, currently the head of the OECD's Indicators and Analysis Division with the Education Directorate, “presented a proposal to rename the BPC to ‘PISA Governing Board’, following a new practice progressively introduced in other areas at the OECD” (OECD/EDU 2004:3). For some members, this was a relatively innocuous development; for others, it was seen as a strategic move by the Education Directorate to exert more control over the Board. The BPC was originally created to “run differently from the rest of the OECD” (Interview Respondent #6, 5 June 2006). It was created as an autonomous board so as to function in a more decentralized fashion while providing the U.S. Department of Education with some control over the project.

When, in March 2004, the BPC was renamed and became the PISA Governing Board (PGB), BPC members were under the impression that this renaming “would not have implications on the role, mandate and operation of the Board” (OECD/EDU 2004:4). It nonetheless provided a mechanism for the Education Directorate to exert more control over the PISA Governing Board and thus to achieve OECD-related organizational objectives. This shift in power
relations also resulted in less influence by the U.S. Department of Education over the trajectory of the OECD PISA.

Another change in power relations occurred in the relationship between the Indicators of Education Systems (INES) Project and the PISA Governing Board. Originally, Network A, which had created the OECD-based assessment, had overall responsibility for the study. The Board of Participating Countries was established with the mandate to oversee the assessment. The relationship of the BPC to Network A was depicted in 1997 in the Data Strategy’s ‘Proposed Project Structure and Management’ as:

Figure 7: The Relationship of the BPC to Network A (1997)

By 2005, this relationship had altered to one where Network A played a significantly smaller role in the OECD PISA’s overall development. The Education Directorate gained more control over the governance of the OECD PISA. When the Board of Participating Countries (BPC) was transformed into the PISA Governing Board (PGB), it signified a more senior level of policy involvement. The delegates within the Board of Participating Countries (BPC) were gradually replaced with more senior level policymakers.
At a Network A meeting in 2005, Education Directorate representatives informed Network A delegates that it was time to let go of their control over the OECD PISA project. Network A members had originally assigned themselves the responsibility for the future development of the OECD PISA. The PISA Governing Board had, however, created its own sub-committee responsible for this role – the PISA Strategic Development Group. The members of Network A were told that even though Network A had originally created the OECD PISA, they could no longer exert any influence over it (Interview Respondent #19, 24 September 2006).

The organizational relationship between the PISA Governing Board and Network A is represented today as:

---

**Figure 8: The Relationship of Network A to PGB (present)**

---

In October 2006, the U.S. Department of Education informed INES Project’s Network A and the Education Directorate that they would be stepping down as Chair and consequently withdrawing full funding for the network. Funding was to be reduced from 90 to 70 to 50% from 2006 to 2009. The U.S. Department of Education proposed that member states equally share funding for
the network which amounted to $400,000 yearly (OECD/INES 2006). It appears that under George W. Bush's Administration, INES Project Network A was no longer a strategic objective that addressed the Department of Education's domestic needs.

Through its institutional structures and administrative processes, the OECD exerted control over PISA work and pursued its objectives as a policy actor. By instituting certain procedures and standards, it was able to govern the conduct of members involved in the production of PISA indicators. For example, it planned regular meeting schedules for the National Project Managers, the Expert Groups and the PISA Governing Board. The Education Directorate produced yearly budgets and financial statements. It edited and published all OECD PISA publications. It circulated meeting minutes and other documentation to ensure communication took place among all group members.

**Figure 9: Organizational Components of the OECD PISA**
Relations of domination ensured agreement took place, particularly inside a political organization such as the OECD where stakes were high for participating states and considerable financial contributions were being made. For example, the PISA Governing Board (PGB) is composed of members representing their respective ministries of education. Even though during meetings, members are allowed to speak on behalf of their state interests, these interests have to be aligned with those of the Education Directorate. Under certain conditions, the OECD Secretariat can exert pressure so that a delegate does not block OECD work. As one of the interview respondents noted:

The OECD Secretariat has objectives and goals to achieve and knows how to achieve them. They lean on country delegates in order to achieve these objectives. There have been incidents in the past whereby the OECD Secretariat and successive directors have complained to ministers or top officials when a country delegate is seen as hindering the attainment of these objectives or as being obstructive (Interview Respondent #16, 30 August 2006).

The OECD's institutional structure is a powerful force for disciplining participation in the PISA and for regulating the “rights and duties of the countries engaged in research” (Bottani and Vrignaud 2005: 77). A cost structure is in place that monitors contributions by each participating country. The contributions are established according to an OECD contributions scale. This is a rigid structure to which participating states are held accountable for their payments to the OECD PISA’s organizational structure. Contributions made by participating states in the PISA are “based on the general scale of country contributions to the OECD” (OECD/EDU 2003).
Because 60% of OECD PISA funding comes from the United States, Japan and Germany, often times they are able to influence the final decisions made on committees (Interview Respondent #17, 18 September 2006). For example, the U.S. Department of Education contributes 25% to OECD PISA's international overhead costs. The Japanese National Institute for Educational Policy Research is the second largest financier and the third largest is the German Federal Ministry of Education and Research\(^8^2\) (OECD/EDU 2003). The American and the Japanese delegates can at times exert more control by virtue of being the major contributors to the OECD PISA. At other times, these actors need to compromise.

The Education Directorate publicizes the reliability and the validity of the PISA data and their importance as a measure of educational performance through its publications. The OECD has not limited itself simply to providing educational indicators. It has taken the role of analyzing the indicators so as to provide policy advice. This is a more evaluative and normative role for the Education Directorate.

The OECD Secretariat reorganizes and restructures its directorates so as to ensure its capacity for official knowledge creation and knowledge production is optimized. As a policy actor, the OECD Secretariat is concerned with maintaining its relevance in what it views as a "new world context" (OECD 2004e).

\(^{82}\) The percentages are based on the OECD's 2003 scale.
Even though the Education Directorate and the PISA Governing Board exert control over the knowledge produced by the OECD PISA, they have to meet the objectives assigned to them by more senior officials. In 2003, the Directorate of Education held its first ‘Chief Executives of Education’ meeting. The Chief Executives came together to discuss the future dilemmas they faced in education. Out of these discussions, five themes emerged that focused on educational outcomes and quality as well as social cohesion and societal learning.

At their second meeting in 2005, the Chief Executives requested that the OECD PISA’s work “further evolve in order to provide better insights into the causes and consequences of educational outcomes as well as issues of efficiency and cost effectiveness in education, rather than just repeatedly document[ing] the levels and distributions of outcomes” (OECD/EDU 2005b:7). The original problematization of the PISA work was the provision of regular data on the cumulative yield of educational systems. The Chief Executives were now demanding more information from instruments that were originally designed to retrieve indicator data on student achievement.

The Chief Executives pointed to the limitations of the OECD PISA, noting the following: “The contribution of PISA to our understanding of how our educational systems worked was important but that, while hard facts are

---

83 The label ‘Chief Executives of Education’ was invented by the OECD Education Directorate to encompass the diversity of titles that existed for those senior government representatives responsible for education such as ministers of education or directors of education (Interview Respondent #19, 24 September 2006).
important for identifying problems, soft facts were more likely to lead us to the solutions" (OECD/EDU 2005b: 5).

The quantitative results produced by the OECD PISA provide useful indicators, but in order to understand more deeply what was going on, the Education Chief Executives felt that qualitative analyses were required. They referred to a recent review of the Danish educational system as an example. This review was similar to past OECD case study reviews, but this time, it applied OECD PISA indicators as a diagnostic tool by identifying what the problems and successes were for Danish students. In this type of country review process, OECD PISA results were "integrated" into the "expert-based peer review methodology" (OECD 2005b: 11). A group of reviewers was sent in to a country to conduct a qualitative assessment and present its core findings in a report. The Education Directorate responded to this request for more national reviews and initiated several reviews in 2005 which focused on equity of educational systems. The Swedish, Spanish and Finnish education systems were reviewed.

The review of the Finnish education system, for example, discussed the causes for the high level of equity in the Finnish education system. One was the presence of a strong welfare state which "prevents certain barriers to education, like chronic bad health, or housing shortages that require families to move consistently and create instability in schooling" (OECD 2005d: 16). The report discussed how Finnish educators could draw on the assistance of a wider network of professionals. These processes ensured students who were faced
with non-school related problems at home or in their communities could get the help they needed from social workers, counselors, and psychologists. The report found that the “involvement of the strong Finnish welfare state is crucial to the success of education.” In particular, the authors noted, “The Finns take it as obvious that both high-quality schooling and co-ordination with non-school services are necessary for equity” (21). Even though the review did initially use OECD PISA indicators to help map what to look for, it shifted the focus away from the statistical approach of the OECD PISA and strengthened a more in-depth qualitative approach to educational research.

The example of the Finnish report demonstrates that the OECD could be a venue not only for an evaluative and statistical method for assessing school performance but also for a more informed and nuanced qualitative methodology. It also demonstrates how the PISA can be used to open up new areas of discourse and allow for more debate on the types of policy formulations that could be prescribed. It helps shed light on how certain aspects of the neoliberal project, particularly when it is combined with Third Wayism, may serve as entry points for expanding social democratic policies (Levitas 1998). As I have argued, the PISA is constructed as a scientific attempt at measuring and evaluating student performance and at exerting control over the knowledge schools produce. Yet this entity is fragile since it is conceived as part of a neoliberal project that is constantly mutating and adapting to changing political, economic, and social realities.
At the Chief Executives of Education meeting held in Copenhagen, Denmark from September 22 to 23, 2005, the Deputy-Secretary General of the OECD, Berglind Ásgeirsdóttir, emphasized that OECD work has to be relevant to the needs of its members (OECD/EDU 2005b). She noted that the other OECD education programs had received less attention than the OECD PISA even though their work was just as relevant to member countries and "deserves to be well known too." The Deputy-Secretary General, who is from Iceland, pointed to other areas of concern to ministries of education such as "out-of-school youth, vocational pathways into higher education, who should pay for higher education, the importance of competencies, changing skills sets and ageing" (OECD/EDU 2005b: 5). Because of these competing priorities that the Education Directorate faces, it cannot sustain as much focus on the knowledge production of educational statistics. Interest in educational statistics cannot be sustained and financing of educational monitoring has to be considered along with other areas. These realities contribute to the PISA's fragility since the science of educational measurement cannot replace all other sources of educational knowledge.

In contrast, as we shall see in the next section, the International Association for the Evaluation of Student Achievement's (IEA) work is devoted to the assessment of students. Because the IEA is committed to one area of work, that of educational measurement, I suggest that it can exert more control over knowledge production processes. Furthermore, the academic grounding of

84 Prior to her position at the OECD, Ásgeirsdóttir was Secretary-General in the Ministry of Social Affairs in Iceland since 1999 and between 1988 to 1996.
its community members builds relations of power in the realm of the educational measurement discipline.

**The International Association for the Evaluation of Student Achievement**

In the mid-1990s, the IEA was perceived as being a weak organization that lacked the financial resources and the organizational capacity effectively to implement large scale international assessments. It has since built up its organizational capacity and instituted more rigid quality standards that have strengthened the quality of its studies.

The IEA decided to stabilize the cycle of testing, and conducted another TIMSS in 1999 (now known as the Trends in Mathematics and Science Study). Since 1999, there has been TIMSS 2003, and TIMSS 2007 is in the planning phase. The IEA has also experimented with new techniques for understanding teaching and learning, such as its video study of teachers teaching in the classroom. In 2001, a decade after the IEA Literacy Study, the IEA initiated regular surveys of the Progress in Reading Literacy Study (PIRLS). As of 1999, IEA studies have been funded by the U.S. National Centre for Educational Statistics, the U.S. National Science Foundation, and the World Bank. It is important to note that the World Bank has taken an interest in educational indicators since it can monitor the performance of educational systems that it is funding.

The IEA studies are not limited to assessing students in mathematics, science and reading. Other areas of study include how information and
communication technologies are implemented and used in educational systems and classrooms. The first Computers in Education Study (COMED) took place in 1989 and 1992 and the Second Information Technology in Education Study (SITES) consisted of two modules – the first was implemented in 1997-1999 and the second in 2000-2002. The Civic Education Study (CIVED) was another study that the IEA implemented in two phases: 1996-1997 and again in 1999 and 2000 (Kankaanranta 2005).

Even though the IEA was portrayed by several interview respondents as weaker than the OECD, as having less "clout", and as having less impact on the policy debate as compared with the OECD PISA, the IEA remains a strong and respected organization among members of the international student assessment community. One of the interview respondents described the difference between the IEA and the OECD PISA in the following manner: “The IEA is doing things that are far more germane to education. Let me put it this way – PISA is an accountability engine. It tells you hardly anything about teaching and learning. It tells you that there is a problem but it doesn’t tell you how to fix it” (Interview Respondent #7, 9 June 2006).

The IEA has opted not to become involved with the OECD PISA. Even though for each OECD PISA cycle, a tendering process is in place, the IEA does not participate in it. At the same time, the PISA Governing Board (PGB) finds it problematic that there have been very few proposals for the implementation of the OECD PISA. ACER has been the prime contractor for OECD PISA 2000, PISA 2003 and PISA 2006. ACER faced little competition.
The PGB tried to increase the number of tenders it receives. It was worried that ACER, the leader of the International Consortium, had gained too much power and control over the implementation of the OECD PISA. It thus encouraged members of the original ACER-led consortium to submit separate proposals. For example, for the PISA 2009 call for tender, Citogroep split from ACER and formed its own consortium (Interview Respondent #14, 27 August 2006). This subdivision of the ACER-led consortium helps build European capacity for educational measurement since Citogroep is a Dutch research agency.

The PGB has gone so far as to request that strategies be taken to "stimulate the market" so that there are more researchers and research agencies submitting proposals in response to future calls for tender. These strategies include technical seminars for developers, changing the contract structure so that it is more of a multi-contract allowing for small contractors to bid for smaller jobs. The contract process was changed into two stages. The first stage gathered a large number of initial proposals in response to the call for tenders. The second stage was a more restricted submission of proposals from the pool selected during the first stage (OECD/DEELSA 2003: 5).

The challenge many ministries of education face is in finding the financial resources to implement student assessments conducted by both the IEA and the OECD PISA. For some this is quite expensive. The French Ministry of Education, for example, has opted not to participate in TIMSS 2006 but only in PISA 2006 because they could not afford two assessments (Interview
Respondent #16, 30 August 2006). In contrast, the German Federal Ministry of Education and Research is participating in both TIMSS 2006 and PISA 2006 assessments.

The IEA considers that in order for the OECD PISA and the IEA to “achieve harmony and cooperation the two organizations have to be regarded as equal partners.” There have been recent attempts made to initiate a process of collaboration, since states involved in both programs want them brought “closer together” (IEA 2005). Since the IEA studies test nine year-olds, an area of collaboration could involve the IEA implementing an OECD PISA on a younger cohort.

Network A of the Indicators of Education Systems (INES) Project, a technical forum, has acted as an intermediary between these two entities so that discussions could begin for such an initiative. The first rapprochement took place in March 2005 at a Network A meeting when Seamus Hegarty, the IEA chair, publicly addressed Network A delegates. The second encounter also took place at a Network A meeting in July 2005. At this meeting, Hegarty suggested that a “high-level standing group of the key players in international comparative education studies” be created so as to “exchange information” (OECD/INES 2005: 8). So far, nothing has transpired from this suggestion.

In October 2005, several representatives from Network A attended the IEA’s 46th General Assembly. They shared a statement with the Assembly that was agreed upon by the members of Network A. This statement encouraged the IEA and the OECD PISA to continue talking to each other in order to achieve
“understanding among and between the international agencies” (IEA 2005). Following the statement, the IEA’s General Assembly issued a motion that a “cooperative task-force be formed with members from both the IEA General Assembly and the PISA Governing Board (PGB)”. The task force’s mandate was to “ensure cooperation between the IEA and the OECD in order to foster the continued quality and utility of international comparative studies for member countries” (IEA 2005).

When the IEA approached the OECD PISA Governing Board in order to discuss the establishment of this task force, it was handed a draft Memorandum of Understanding that established a collaborative relationship between UNESCO and the OECD to pursue an assessment of 9 year olds (Interview Respondent #22, 31 October 2006). From the vantage point of the IEA, this was another political maneuver that demonstrated that the OECD PISA was not a trustworthy partner. Even if the IEA and the OECD PISA were to reach an agreement on working together on an assessment, one of these organizations has to make a compromise. As one of the interview respondents put it, “If these two tests are brought together, the issue becomes which organization will take control and which organization will give up control” (Interview Respondent #12, 15 August 2006).

The U.S. Department of Education continues to fund and support the IEA studies (Interview Respondent #19, 24 September 2006). The IEA studies are useful for making comparisons with the National Assessment of Educational
Both NAEP and TIMSS test fourth and eighth grade students and share similar frameworks (Neidorf, Binkley, Gattis, and Nohara 2006). In combination, NAEP and TIMSS provide the Department of Education with a potent monitoring tool, one that is more instrumental to their evaluation needs than the OECD-based assessment. As I noted earlier, this is especially useful with the TIMSS implementing a scale ranking method that is constructed in terms of international benchmarks.

Even if the OECD attempts further to increase its reach in the international assessment field, it cannot eliminate the IEA as a major player. Ministries of education find it useful to have an assessment that is curriculum-based and with which they can compare the results of their national assessments. They prefer not to depend solely on an OECD-based student assessment as an instrument for measuring their students' performance, particularly since it is a non-curriculum based assessment.

Several members of the PISA Governing Board (PGB) have requested at various times during the OECD PISA's survey cycles the inclusion of a younger age cohort in the assessment. This can be interpreted as an attempt to venture into IEA 'territory', since the IEA studies test 9 and 13 year old students. Several respondents have indicated that the OECD Secretariat's Education Directorate is interested in dominating the field of international assessment and in eventually eliminating any competitors, including the IEA (Interview Respondent #7, 9 June 2006; #16, 30 August 2006; #17, 18 September 2006).

\[^{85}\] Also known as the Nation's Report Card.
This evidence suggests that the OECD, as a policy actor that increasingly behaves as an "international consultancy," is in fact trying to "enlarge its markets" (Marcussen 2004: 100).

The initial problematization of an OECD-based assessment was justified in terms of the weaknesses of the IEA studies. These studies were criticized for their poor quality and for not being regular and timely. The IEA was criticized for its poor management capacities. The IEA studies were portrayed as weak because of their academic orientation and an OECD-based assessment was portrayed as strong because of its policy orientation. Yet it is precisely because of the IEA's concern for academic based research that today it derives its strength. It can generate from the vantage point of the discipline of international educational measurement the official knowledge that educational policymakers and researchers rely on.

Because of the IEA's academic orientation, it can expand its studies of student learning and teaching beyond the economic context of education. As the IEA's Executive Director, Hans Wagemaker, has indicated, "it is reasonable to expect that the work of IEA should contribute to our understanding of teaching and learning, to the educational policy debate within countries and internationally, and, ideally, to research practice itself" (Wagemaker 2004). The IEA offers educational researchers more opportunities for secondary research analysis, for broadening research knowledge, and for meeting with individuals who are like-minded.
The IEA has instituted a venue for researchers to share their secondary research findings. The first International Research Conference on Student Achievement was held in Cyprus in May 2004. The 2nd conference took place in November 2006 and was organized in collaboration with the Brookings Institution. The conferences were free, which likely attracted wider participation. During these conferences, academics met to present and discuss research they conducted on results from IEA studies. In contrast, the OECD PISA does not have a venue for establishing a research dialogue among members of the academic community. Even though forums in mathematics and sciences for discussing framework concerns were instituted, these mechanisms were created to serve political goals rather than to fulfill academic research objectives. Other meetings that the OECD organizes are concerned with sharing 'success' strategies implemented by high performing OECD PISA countries, such as Finland. Their target audiences are policymakers and educational administrators rather than academics. If the OECD wants to maintain itself as a legitimate policy actor in the realm of international educational assessments, it would have to bolster its technical capacity continuously.

The IEA is not a weak player in the international assessment arena, but has continued to strengthen itself. It has established collaborative relationships with major international organizations such as the World Bank and the Inter-American Development Bank, the European Commission, and UNESCO. Furthermore, the IEA has collaborative relationships with regional assessment organizations such as the Southern Africa Consortium for Monitoring
Educational Quality (SACMEQ) project, the Programme d'Analyse des Systèmes Educatifs des Pays de la CONFEMEN (PASEC), and the West African Examinations Council (WAEC) (Interview Respondent #22, 31 October 2006). In effect, the IEA is an important player within the global architecture of education.

The IEA trains researchers not only in the technical aspects of IEA studies but also in the way the data are used in policy analyses. Future training will focus on "assisting researchers to address more fundamental questions, such as what can (and should be) be answered by data that are collected from IEA studies and how that data can be integrated with other policy relevant information" (Wagemaker 2004). The IEA has improved its dissemination and communications strategy, ensuring that its publications and data are widely distributed, it gets the press coverage it deserves, and it builds public awareness on the assessment work it conducts. It has also developed a user-friendly website for its studies so that publications, data, media releases and other information are easily accessible both to researchers and to the general public. Recently, the IEA has collaborated with the Educational Testing Service to form the IEA-ETS Research Institute (IERI). The Research and Development Division at ETS and the IEA Data Processing Center will work together to advance statistical and psychometric methodologies and knowledge and to share advances in the educational measurement field. From a public relations perspective, the IEA may not have the clout or the vast resources that the
OECD has, but it is effective in maintaining the relevance of its studies to educational researchers and to participating countries.

The IEA has a stronger hold on the disciplinary knowledge of the field of international student assessment than the OECD. The IEA founded the field of international student assessment. It brought together leading experts from various disciplines who were committed to its mission and often worked on a voluntary basis. Its members pioneered many of the techniques and methods currently applied in international assessments. The issues that the community members of the PISA debated in the 1990s had already been discussed and debated by those involved in the IEA in the 1960s and the 1970s. Issues such as whether to use open-ended test items or multiple choice items; whether to define the population in terms of age or grade; and the techniques for data analyses had been discussed thirty years earlier by educational researchers and psychometricians involved in the IEA studies (Wolf 2004).

Conclusion
In this chapter, I have discussed the relationships of power that characterize the formation of the power bloc. I argued that power relations can be traced by examining how international student assessments attempt to govern the behavior of educational actors such as students, teachers and schools. I examined the way in which the literacy object was constructed by the PISA instruments. I suggested that by creating an instrument for assessing the literacy of students, a curriculum template is imposed on educational systems.
This new curriculum would create the knowledge workers destined for the global economy. The second section of the chapter was concerned with analyzing the relationships of power that exist within institutional structures. I concluded that there are tensions that arise as the OECD tries to compromise among its multiple roles of policy forum, policy instrument and policy actor. At the same time, the OECD is concerned with maintaining its relevance in an increasingly competitive environment where several organizations and institutions generate policy ideas and solutions. I also discussed how the IEA has enhanced its power relations by collaborating with other organizations and by continuing to expand its technical capacity.

In the next chapter, I will examine the global relations of power and revisit the theme of the fragility of the PISA.
Chapter 9
Global Architecture of Education and Fragility

Introduction

In Chapter 8, I argued that the construction of an object called ‘literacy’ helped govern the conduct of schools by setting standards for a new curriculum – one that created the global knowledge worker. I discussed the institutional power relations that exist within the OECD and the IEA. In this chapter, I return to two themes that are important to my analysis – that of the global architecture of education and the fragility of the PISA as a power bloc.

The OECD PISA and the knowledge it produces interface with an international network of users and agents. In earlier chapters, I mentioned various agents that the IEA and the OECD PISA coordinated with. I would like to discuss in more detail how the OECD PISA operates within the global architecture of education. The analysis includes a discussion on the profit making educational assessment industry.

The second theme discusses the fragility of the PISA and points to how technical capacity can be undermined in the field of international student assessment. As we saw in Chapter 3, the history of constructing an object called ‘intelligence’ was riddled with controversies and debates. We also saw in Chapter 4 that the field of international student assessment had its own controversies. Some of these difficulties return to haunt the Programme for International Student Assessment (PISA).
The PISA within the Global Architecture of Education

Within the global architecture of education, the PISA has various functions. Its results are used by end-users who view it as a legitimate conveyer of official knowledge. For ministries of education, it measures their student performance as knowledge workers. For policymakers, the PISA is a tool for measuring how well schools are doing. For the media, the PISA becomes another news item to report on. For the Education Directorate, the PISA is a data producing instrument that feeds into its knowledge production capabilities. For the World Bank, the PISA data are tools for monitoring educational quality in recipient states.

The official knowledge produced by the PISA is destined for the end user. The users can be divided into several groups. At the top of the knowledge-seeking hierarchy are experts who conduct complex statistical analyses using the PISA data. At the bottom is the general public with little knowledge of statistical models and for whom the results are reported in terms of league tables and rankings. Policymakers interact with the PISA when results are reported to them in international or national reports. They are propelled into action when students do poorly on international assessments. Educational researchers with some statistical knowledge can download the PISA results from the database and conduct their own secondary analyses.

The general public interacts with the PISA when results are reported in the media as rankings or league tables. This form of inscription device is effective in raising public debate over educational reforms and in arousing a
political response. It is the most generalized and aggregate form of representation of student performance. A number such as 546 or 448 tells very little about student learning. Yet when it is ranked against another number and associated with a particular national grouping, it becomes very potent.

Figure 10: An example from an OECD PISA publication reporting PISA 2000 results

<table>
<thead>
<tr>
<th>Reading performance in PISA 2000</th>
<th>Percentage at:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 1 or</td>
</tr>
<tr>
<td></td>
<td>Level 5</td>
</tr>
<tr>
<td>Mean score</td>
<td>below</td>
</tr>
<tr>
<td>Finland</td>
<td>546</td>
</tr>
<tr>
<td>Canada</td>
<td>534</td>
</tr>
<tr>
<td>New Zealand</td>
<td>529</td>
</tr>
<tr>
<td>Australia</td>
<td>528</td>
</tr>
<tr>
<td>Ireland</td>
<td>527</td>
</tr>
<tr>
<td>Korea</td>
<td>526</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>523</td>
</tr>
<tr>
<td>Japan</td>
<td>522</td>
</tr>
<tr>
<td>Sweden</td>
<td>516</td>
</tr>
<tr>
<td>Austria</td>
<td>507</td>
</tr>
<tr>
<td>Belgium</td>
<td>507</td>
</tr>
<tr>
<td>Iceland</td>
<td>507</td>
</tr>
<tr>
<td>Norway</td>
<td>505</td>
</tr>
<tr>
<td>France</td>
<td>505</td>
</tr>
<tr>
<td>United States</td>
<td>504</td>
</tr>
<tr>
<td>Mexico</td>
<td>497</td>
</tr>
<tr>
<td>Switzerland</td>
<td>494</td>
</tr>
<tr>
<td>Spain</td>
<td>493</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>492</td>
</tr>
<tr>
<td>Italy</td>
<td>487</td>
</tr>
<tr>
<td>Germany</td>
<td>484</td>
</tr>
<tr>
<td>Hungary</td>
<td>480</td>
</tr>
<tr>
<td>Poland</td>
<td>479</td>
</tr>
<tr>
<td>Greece</td>
<td>474</td>
</tr>
<tr>
<td>Portugal</td>
<td>470</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>441</td>
</tr>
<tr>
<td>Mexico</td>
<td>422</td>
</tr>
</tbody>
</table>

| Finland                        | 7.0            |
| Canada                         | 9.9            |
| New Zealand                    | 13.9           |
| Australia                      | 12.5           |
| Ireland                        | 11.0           |
| Korea                          | 5.8            |
| United Kingdom                 | 12.9           |
| Japan                          | 10.1           |
| Sweden                         | 12.6           |
| Austria                        | 14.6           |
| Belgium                        | 19.0           |
| Iceland                        | 14.5           |
| Norway                         | 17.5           |
| France                         | 15.2           |
| United States                  | 17.9           |
| Mexico                         | 17.9           |

| Finland                        | 18.5           |
| Canada                         | 16.8           |
| New Zealand                    | 18.7           |
| Australia                      | 17.6           |
| Ireland                        | 14.2           |
| Korea                          | 8.7            |
| United Kingdom                 | 15.6           |
| Japan                          | 9.2            |
| Sweden                         | 11.2           |
| Austria                        | 8.8            |
| Belgium                        | 12.0           |
| Iceland                        | 8.1            |
| Norway                         | 11.2           |
| France                         | 8.5            |
| United States                  | 12.2           |
| Mexico                         | 8.2            |

Source: OECD 2004c: 5

When OECD PISA results are reported nationally, they have different effects, particularly for OECD member states. For example, the German public,
policymakers and politicians were shocked with the performance of their students in PISA 2000. In fact, the results of the OECD PISA were referred to by the Education Ministry as the "PISA shock." German students ranked below the OECD average. The PISA results pointed to the inequity in the German educational system, where a large proportion of immigrant children were "under-achievers." Such results were viewed as shameful. German education ministers and politicians reacted by instituting reforms and unifying standards across the 16 Länder (Schmid 2003; Williamson 2003). The PISA results instigated recommendations that neoliberal reforms be instituted in the German educational system (Allmendigner and Leibfried 2003).

The media reported the results of Japanese students' performance on PISA 2003 in the following manner:

The reading skills of Japanese 15-year-olds have plummeted over the past few years, leaving Japan lagging behind 13 other states, a survey by the Organization for Economic Cooperation and Development (OECD) has shown.

The OECD's Programme for International Student Assessment ranked Japan at 14th place in students' ability to read and comprehend graphs and written passages in the 2003 survey, whose results were released Tuesday. In the first survey conducted in 2000, Japan had ranked eighth in this category.

The same newspaper article reported that "In response to the results of the survey, the ministry plans to introduce a program to boost students' reading comprehension skills by next summer" ("Japanese kids lag world in reading skills" 2004). Policymakers did not further investigate the results. They trusted the numbers and set out to reform the educational system. The rankings in this
particular cultural context propelled policymakers to act in response to these numbers.

In other jurisdictions, the OECD PISA results were used to bolster the platform of special interest groups. For example, the PISA 2003 results of New Zealander students pointed to a decline in performance. A special interest group used the test results to advocate the marketisation of education and school choice as the best way to improve student achievement (Ormond 2004). American businesses recently used international assessment data as evidence for the need to reform American schools so as to produce more students interested in mathematics and engineering (Business Roundtable 2005). In fact, the Wall Street Journal reported that the poor performance of students in mathematics was an "economic time bomb" and "trailing other OECD countries on education measures may reduce U.S. economic growth by as much as a half percentage point a year" (Kronholz 2004).

Academic researchers have used the PISA data to support their arguments for social democratic policies rather than narrowly conceived neoliberal ones. For example, Berliner (2005) has drawn on disaggregated international student assessment results as evidence of the linkages that exist between poverty and poor student performance on international tests. Ethnicity is used as a proxy for poverty since there are high inter-correlations of poverty, ethnicity and school achievement in American schools. The disaggregated data depicted a snapshot of reality that was different from that of the OECD PISA's international rankings: White American students ranked near the top and
African American and Hispanic students ranked at the bottom internationally.

Rankings can be powerful tools not only for legitimizing market driven policies but for advocating for social change.

**Figure 11: Sample of student performance rankings with American populations disaggregated by ethnic grouping**

<table>
<thead>
<tr>
<th>Country</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>557</td>
</tr>
<tr>
<td>Korea, Republic of</td>
<td>547</td>
</tr>
<tr>
<td>New Zealand</td>
<td>537</td>
</tr>
<tr>
<td>Finland</td>
<td>536</td>
</tr>
<tr>
<td>Australia</td>
<td>533</td>
</tr>
<tr>
<td>Canada</td>
<td>533</td>
</tr>
<tr>
<td>United States Average Score for White Students</td>
<td>530</td>
</tr>
<tr>
<td>Switzerland</td>
<td>529</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>529</td>
</tr>
<tr>
<td>Belgium</td>
<td>520</td>
</tr>
<tr>
<td>France</td>
<td>517</td>
</tr>
<tr>
<td>Austria</td>
<td>515</td>
</tr>
<tr>
<td>Denmark</td>
<td>514</td>
</tr>
<tr>
<td>Iceland</td>
<td>514</td>
</tr>
<tr>
<td>Sweden</td>
<td>510</td>
</tr>
<tr>
<td>Ireland</td>
<td>503</td>
</tr>
<tr>
<td>Norway</td>
<td>499</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>498</td>
</tr>
<tr>
<td>United States Average Score</td>
<td>493</td>
</tr>
<tr>
<td>Germany</td>
<td>490</td>
</tr>
<tr>
<td>Hungary</td>
<td>488</td>
</tr>
<tr>
<td>Spain</td>
<td>476</td>
</tr>
<tr>
<td>Poland</td>
<td>470</td>
</tr>
<tr>
<td>Italy</td>
<td>457</td>
</tr>
<tr>
<td>Portugal</td>
<td>454</td>
</tr>
<tr>
<td>Greece</td>
<td>447</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>446</td>
</tr>
<tr>
<td>United States Average Score for Hispanic Students</td>
<td>437</td>
</tr>
<tr>
<td>United States Average Score for African American Students</td>
<td>423</td>
</tr>
<tr>
<td>Mexico</td>
<td>387</td>
</tr>
</tbody>
</table>

Source: Berliner 2005: 20

The PISA is an important node in the global architecture of education, particularly for the World Bank. The World Bank's Independent Evaluation Group (IEG) assesses the Bank's programs and policies and reports directly to the Bank's Board of Executive Directors. It uses educational statistics to
measure the effectiveness of the Bank’s financing in the area of education. For example, in its reports on education loans made to the Romanian and Peruvian governments, the IEG used the PISA data (and data from IEA studies) to assess whether improvements have been made in learning outcomes and in educational quality (Berryman, Gove, Sapatoru, and Tirca 2007; Benavides, Carnoy, Cueto, and Gove 2007). These educational statistics are cost-effective research tools that address the Bank’s work practices. The Bank’s research agenda has increasingly been to provide evidence of efficiency gains made through educational reforms (Mundy 2002).

Another significant aspect of the global architecture of education is the marketization of educational assessment products, services and know-how. Participants in this multi-million dollar industry supply educational measurements services and products to national educational systems, regional assessment programs and international studies of student achievement. For example, ACER in 1998 earned $15 million in revenue and had only a few international contracts. In 2005, its revenue had increased to $36 million and it had more than 15 international contracts worldwide, including the OECD PISA (ACER 1999; ACER 2005).

Educational measurement centres located in OECD member states compete with one another for contracts. Industrialized, newly industrialized and oil-rich economies send out requests for proposals for specific educational measurement services. For example, Ontario’s Ministry of Education hired ACER to develop reading items for its Ontario Secondary School Literacy Test.
Some contracts are awarded through international and regional organizations such as the World Bank or the European Union. For example, the World Bank funds the institutionalization of measurement educational techniques in less developed economies. Informational and data-gathering systems for assessing and testing students are set up in these states. Other contracts are awarded by government entities interested in bilateral aid (Cito 2006).

The European Union has also been active in educational funding. In 1989, it created a program called PHARE (Pologne, Hongrie Assistance à la Reconstruction Economique) which was originally set up to assist Central and Eastern European states in their transition from a communist system to a liberal economic one. Over time, PHARE has become the “EU's main financial instrument in support of the accession process” (Martens 2000: 2). From 1996 to 1999, the EU/PHARE program funded Poland’s implementation of a national assessment program for its educational system. In 1999, EU/PHARE funded a vocational certification and assessment program in Slovenia which became an EU member in 2004 (Cito 2007).

Some research centres are non-profit organizations such as the Australian Council for Educational Research (ACER). Others, such as the Netherlands National Institute for Educational Measurement (Cito), were initially established as public organizations and were subsequently privatized. Cito was part of the ACER led consortium for the OECD PISA 2000, 2003, and 2006. When these organizations win large enough contracts, they open new offices.
For example, ACER has opened an office in the United Arab Emirates where it was contracted to develop assessments for post-secondary education selection and for student achievement (ACER 2005). Cito, which was privatized in 1999, has established German and American subsidiary companies. It also has an office in Turkey where it won a contract for developing a ‘pupil monitoring system’ for students aged 4 to 12 (Cito 2007).

The global market for international student assessments is held together by several information gathering, standardizing and overlapping organizations and entities in which the PISA is one node. The global marketplace is connected to the global architecture of education. The OECD’s Education Directorate is part of the network. The Education Directorate’s Indicators and Analysis Division, which administers and manages the PISA and the INES Project, feeds information into the global architecture. It provides tools for helping to govern at a distance.

One of my interview respondents complained that the OECD PISA monopolizes the market and thus hinders free market competition: “When you put the measures in a political organization, because that’s what the OECD is, you create two kinds of incentives: one - you politicize the data, because its big money, two - you create financial incentives for doing this at the exclusion of other people who would compete” (Interview Respondent #7, 9 June 2006). It appears that some members of the international student assessment community identify themselves as producers of services and products. The respondent was referring to the unfair competitive practices of the Education Directorate in the
globalized market for international student assessments. Thus, the OECD PISA becomes one product in the international marketplace that is being produced and packaged by its 'owner,' a political organization. Another interview respondent has pointed to the negative effects of 'politicization' stating, "[b]ecause the PISA is so politically driven, it has become irreproachable" (Interview Respondent #17, 18 September 2006). These concerns over the politicization and unfair market practices weaken the validity and reliability of the PISA data even though its reputation was meticulously built up through its associations with subject matter experts and statistical and psychometric experts.

Fragility of the PISA

The codification, categorization and data production work of the PISA requires that its data be perceived as of high quality, reliable, valid, legitimate and relevant. It is precisely this measurement work that contributes to the PISA's fragility. This section discusses how subject matter experts, psychometricians, statisticians, markers, students and principals are fitting objects to categories. This work is both visible and invisible and involves various communities of practice. I will describe controversies and debates in several areas including cultural bias, student exclusions from tests, quality of the assessments, expansion of the assessment, over-interpretation of results and the modification of the PISA assessment.

Visible Work
The OECD PISA provides an extensive technical account for each of the assessment cycles. This account is similar to previous technical accounts that the IEA published for its studies. It entails a visible account of measurement and codification work, describing how tools and work processes were deployed.

The Technical Report for PISA 2000 discusses all the technical aspects of implementing an OECD-based assessment. It describes how the instruments were designed including the creation of the test items and the questionnaires. It discusses various kinds of operational procedures conducted for sampling, test administration, quality monitoring and translation of items and questionnaires. The report informs the readers on the way the data were cleaned and processed using statistical techniques such as weighting, variance estimation, and scaling. It points out how the data were adjudicated or 'judged' for their overall quality. It presents the results of this adjudication process for each participating state. The report provides details on how scales were constructed for test items and questionnaires. These scales represent the levels of competencies that the items assessed. Groups of items represented certain competencies for each domain to be tested that had been identified through the process of the construction of domain frameworks. Answers from the questionnaires were manipulated and coded to produce indices of variables. For example, the PISA International Socio-Economic Index of Occupational Status was created based on students' responses on parental occupation (Adams and Wu 2002).
Based on this brief summary, one can see that the Technical Report offers a visible account of the codification and categorization work that the PISA does. One has to be very well versed in educational measurement techniques and methodologies in order to question the technical aspects of the OECD PISA. These technical reports legitimize the official knowledge that is produced by IEA studies and OECD PISA and reassures practitioners, policymakers, and the general public of the high reliability and validity of the science of international student assessment. In the next section, I point to some of the invisible work that these types of projects do and discuss the contingent dimension of all types of standardization work.

Invisible Work

Anomalies arise out of work that involves categorization. Value judgments have to be made when categories do not fit. This is an "inevitable consequence of delegated, large-scale alienated survey research labor" (Bowker and Star 1999: 156). For example, markers are hired and trained to mark the tests. They have answer sheets to guide them in their work. At times, however, the responses for open-ended items do not fit the categories in the answer sheets. Markers are forced to use their own judgement to decide whether an answer is correct or not. Students participating in the OECD PISA are asked to fill out background questionnaires. There are questions asking them to categorize their parents' occupations and levels of education. They have to fit their answers into categories. Subject matter experts who create the domain frameworks have to categorize student learning into specific sets of activities and tasks.
One example of invisible work that I would like to discuss in greater detail is that of test item construction – a work process that is intellectually demanding and time consuming. In Chapter 7, I described the work that the PISA domain frameworks do and pointed to how they act as standardized packages. Standardized packages allow for both abstraction and standardization. I noted that standardized packages were effective in changing local practices because they defined concrete practices. In the case of the PISA, however, they did not always succeed in doing so. Even though national teams tried to construct items according to the framework’s standards, team members lacked a working knowledge of these practices. The result was that national teams submitted items that did not necessarily reflect the literacy construct espoused by the OECD PISA. National teams constructed curriculum based test items because that was what they were accustomed to doing, particularly through their work on national assessments and IEA studies. Standards and rules were not able to change ingrained practices.

The domain frameworks cannot change local practices. To succeed in changing practices, intensive training sessions need to take place so that test developers learn to write new items (Interview Respondent #13, 20 August 2006; #20, 20 October 2006; #21, 25 October 2006). The OECD is not the place for this sort of knowledge transfer since it is an organization concerned with economic growth. The OECD does set up sessions to transfer technical knowledge but this type of knowledge is easier to justify to contributing states. Rather, such learning is more likely to occur in academic settings or among
research institutes where students take courses in educational measurement and researchers are continuously advancing their disciplines. I suggest that this type of knowledge transfer may be done better by the IEA rather than the OECD since the IEA is networked into academic circles. As I discussed in Chapter 8, the IEA's institutional structure helps it build a stronger technical capacity for implementing a science of international student assessment than that of the OECD's.

The OECD PISA's International Consortium and the Expert Groups were faced with a shortage of items. Many of the items submitted by the national teams were rejected because they did not fit the frameworks. In other cases, the national teams submitted very few items. Among the expert group members, some had more experience developing these types of test items while others had to learn to write them. Because there was a shortage of items, members of the domain expert group contacted other subject matter experts from their personal networks to assist them in writing additional items. Informal measures outside the formal organizational structures were taken to increase the number of items available for selection and for review (Interview Respondent #21, 25 October 2006).

As the PISA Technical Report informs us, the items were stored in an item bank and were "regularly audited to ensure that the instruments that were developed satisfied framework specifications" (Adams and Wu 2002: 26). Yet, one can never be sure of the quality of the items stored in an item bank since items are not publicly accessible and cannot be scrutinized by educational
researchers. Only a few of the PISA items were released to the public and usually the best examples were chosen for public consumption. According to the practitioners in the field, such a level of secrecy was necessary to ensure that test items could be reused in future assessments and trends could be monitored.

Even though the International Consortium, and more specifically ACER, applies what is considered to be a high quality psychometric model to the analysis of the items, if the items do not accurately reflect the framework then the validity of the assessment may be questioned. If what the PISA instrument measures as its input is not necessarily the same as what it purports to analyze as its output, then how valid is it at measuring the literacy of students and their preparedness for the world of work? (Interview Respondent #20, 20 October 2006). These types of questions arouse controversies. In the next section, I provide a brief overview of the controversies and debates that destabilize the OECD-based assessment and contribute to its fragility.

Controversies and Debates

In this section, I focus on the controversies and debates that not only weaken the PISA but that also destabilize the science of assessing students internationally. More specifically, tools and practices have been refined so as to address errors in sampling populations, cultural bias in tests, and translation errors. Even though this science has been rendered doable through experimentation and tinkering, comparability remains contested.

Cultural Bias
The French Ministry of Education has criticized the PISA instrument, its item construction process, and the test items for being culturally biased. The Ministry has argued that since the OECD PISA is an international assessment that is largely dominated by 'Anglo-Saxons,' it is methodologically, linguistically and culturally biased. The Ministry posits that the European states need to develop an assessment that caters to their specific educational needs and that addresses their particular linguistic and cultural context (Bonnet 2004). The OECD PISA is criticized for being an assessment program overly concerned with "economic and cost effectiveness factors" rather than "learning and teaching considerations based on national practices" (CISAD 2003: 12).

The French Ministry of Education has proposed an alternative assessment - a European assessment which has a different methodological approach that it claims is culturally and linguistically unbiased. It is also an assessment that is supposed to be culturally authentic (CISAD 2003). The French persuaded other European ministries of education and research institutes to work together on a feasibility study (British, Finnish, and Italian). The study aimed to do the following:

- to identify the conditions under which comparative, culturally unbiased data can be obtained from nationally produced test instruments using indigenous, untranslated, original material deemed by experts to be of

86 The French Ministry of Education has also criticized the IALS for being a culturally biased assessment (Bottani & Vrignaud 2005).
87 The feasibility study was funded by the European Union's SOCRATES program which established a community action program in the field of education. SOCRATES I was from 1994-1999. SOCRATES II began to be implemented on Jan. 1, 2000 until Dec. 31, 2006 (Official Journal of the European Communities L28, 2000) (Available at: http://ec.europa.eu/education/programmes/socrates/decsoc2_en.pdf, last accessed March 4, 2007.)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
comparable levels of difficulty between states and to be testing the same skills at similar levels of difficulty (Bonnet 2004:186).

The results of the feasibility study were deemed to be encouraging. A decision was taken to develop a methodological framework for a culturally unbiased study. The original number of participants was expanded to include other European researchers (Belgium Flemish Community, the Dutch, the Norwegians and the Swedes). The project came to be known as the Culturally Balanced Assessment of Reading (C-BAR).

The researchers working on this project were members of the European Network of Policy Makers for the Evaluation of Education Systems. This group was formed in 1995 under the French presidency of the European Union to "facilitate cooperation between European states in the field of school evaluation and monitoring." Until 2006, the network's coordinator and chair has been Gérard Bonnet. Bonnet was the International Project Coordinator for the C-BAR project. Bonnet also was the French delegate to the PISA Governing Board (PGB) from 1997 to 2006. Another researcher involved with C-BAR was Martine Rémond who was also the French delegate on the Reading Expert Group (European Network of Policy Makers for the Evaluation of Education Systems 2006:1).

This example demonstrates the fragility of the PISA. The validity of the tools that it has created for its community members and users can be questioned and contested. The problem of cultural bias can never be fully deleted. C-BAR is offered as an alternative problematization to the OECD-
based assessment as it is a culturally authentic assessment that can address
the weaknesses of the OECD PISA.

Exclusions
The International Consortium established the rules on allowable exclusions and
how these were to be reported by participating states. The allowable exclusion
rate was set at below 5% for each national sample population. Grounds for
decisions to exclude schools varied. Schools were excluded based on their
geographic location or based on the language spoken at the school. Among
the student populations that the OECD PISA allows to be excluded are
“educable mentally retarded students”, “functionally disabled students”, and
“non-native language speakers” who received less than one year of instruction
in the language of the test. The National Project Managers were required to
complete several sampling forms, one of which was the “Excluded Schools”
form for school exclusions. The schools were required to fill out the excluded
column of the Student Tracking Form for student exclusions (Adams and Wu

In order to minimize the number of exclusions and to render the
assessment more inclusive, a special test booklet was created for PISA 2003.
Entitled ‘Une Heure (UH) booklet,’ the test contained items constructed for
students with special needs88 (OECD/PISA 2005). This booklet was developed
with the assistance of the Centre for Educational Research and Innovation’s

88 The UH booklet contained seven mathematics items, six reading items, eight science items
(CERI) working group on special needs. Participating states opted to implement the booklet at the national level but it was not an international requirement.

In the OECD PISA 2003 Technical Report, a table was included that listed each participating state’s overall reported school and student exclusions (OECD/PISA 2005). Any criticism arising in this area can be duly dealt with by pointing towards the care taken to address and report on allowable exclusions. What remains problematic in the sampling process of such international assessments as the OECD PISA is that schools decide which students to exclude. Since definitions of disability vary across cultures (Reid and Knight 1986), this raises the question as to which students schools opt to exclude. Are they students slotted in special education classes? Are they students who are categorized as mildly mentally retarded? Are they students labeled as learning disabled? If this is the case, then it is crucial to take into account the fact that in certain schools, minority and poor students are disproportionately represented in special education (Blanchett 2006). The discussion of student achievement on the OECD PISA becomes irrelevant since the real issues that need to be addressed are poverty, racism and classism (Blanchett 2006; O'Connor and Fernandez 2006; Reid and Knight 1986).

Quality

There have been controversies over international assessments that are not just concerned with the OECD PISA but also with the IEA studies and the International Adult Literacy Survey (IALS). In particular, there have been two main critics – Sig Prais and Harvey Goldstein. These are prominent academics
in their fields. Prais\textsuperscript{89} is a Senior Fellow with the British National Institute of Economic and Social Research. Harvey Goldstein is a (recently retired) Professor of Social Statistics who has been critical of the Rasch methodology.

Prais (2003) has criticized the OECD PISA in an article published in the \textit{Oxford Review of Education}. The usefulness of a non-curriculum based assessment for schools and educational policymakers was questioned by Prais. He discussed the weaknesses of an age-based population sample, focusing on the difficulty in reaching such a student population since many may have left school, some may be in vocational schools and others may be working. He also analyzed the response rates. In the British case, he found that there was a bias "upwards" of the "recorded average scores" since "those who did not participate must be suspected of being, on the whole, in the lower-attaining groups of the population" (Prais 2003: 153). At the end of his discussion, Prais remarked:

"These reservations, taken together, are sufficiently weighty for it to be unlikely that anything of value for educational policy in the UK can be learnt from the PISA survey" (Prais 2003: 153).

Goldstein's (2004) critique was directed at the methodological weakness of the OECD PISA. This methodology relies on a certain item construction model of Item Response Theory – that of the Rasch model. Goldstein pointed to a "mafia" form of control of the methodological approach among statistical agencies and institutions which involved an "adherence to certain procedures for designing, scaling and analyzing tests" (3). According to Goldstein, knowledge

\textsuperscript{89} Prais has analyzed the weaknesses inherent in the TIMSS and the PISA. He finds fault in the sampling design applied for both assessments, whether age-based or curriculum-based (Prais 2005).
production and creation of this "psychometric subset" was reinforced, disseminated and perpetuated through two journal publications – the first was the "theoretical journal", *Psychometrika*; and the second was the "applications journal", the *Journal of Educational Measurement* (3). In fact, Goldstein criticized the inherent cultural bias found in all international student assessments and recommended the creation of regional forms of assessments that were culturally sensitive.

**Expansion of the OECD-based Assessment**

Participation in the OECD PISA grew dramatically from 41 states in 2003 to 57 in 2006. Of the 57 states, 27 did not belong to the OECD. The central concern pertaining to the participation of non-OECD states is that one instrument cannot measure the wide variability that exists among states. The instrument was constructed as a tool for measuring educational yield in OECD member states. As members of the 'Rich Man's Club,' the educational systems of these states are more comparable than non-member OECD states because they are at the same level of economic development. How can one test instrument accurately assess students from Finland and also accurately assess students from Azerbaijan? (Interview Respondent #18, 22 September 2006)? Because of such variability among populations, there exists a high error of measurement for those high performing states and low performing states (Interview Respondent #17, 18 September 2006). This issue also came up with the IEA Six Subject Study discussed in Chapter 4.
There have been discussions among the delegates of the PISA Governing Board (PGB) to adjust the instrument to the needs of the non-member OECD states. The Technical Advisory Group (TAG) has raised this issue with the Education Directorate and with the PGB. The technical experts were concerned that the scale used for the PISA instrument cannot be adequately applied to all states. (Interview Respondent #17, 18 September 2006). The PISA Governing Board delegates have opted against modifying the assessment. The policymakers of the OECD member states designed this assessment to test their students and they were not interested in changing it.

Non-member OECD states have raised the issue that the PISA has to be modified to address their needs. Complaints have pertained to the PISA instrument not reflecting their own culture, particularly since only OECD member states can submit test items. Mexico has often represented the views of non-member OECD countries on PISA Governing Board meetings (Interview Respondent #2, 21 October 2005).90

Over-interpretation of the OECD PISA results

Over-interpretation of the OECD PISA results is a controversial issue for those involved with PISA codification work (Interview Respondents #6, 5 June 2006; #17, 18 September 2006; #19, 24 September 2006). Over-interpretation means that policymakers and politicians make their decisions based solely on the PISA data. Statisticians and educational researchers prefer that policymakers

90 Salzman and Terracino (2006) note in their research work that “during the MAI (Multilateral Agreement on Investment) negotiations Mexico would often claim to speak on behalf of developing countries, apparently to the intense irritation of the other Member countries” (399, footnote 215).
examine more closely the OECD PISA results and conduct secondary analyses before initiating educational reforms. As one of the PISA community members informed me, policymakers expect so much out of the data, "as if the PISA is a magic tool that can tell you everything about an educational system" (Interview Respondent #17, 18 September 2006). According to another interview respondent, policymakers tend to forget that the OECD PISA assessment is a kind of "thermometer," it can detect a problem but it cannot give answers on how to solve the problem (Interview Respondent #6, 5 June 2006).

Statisticians and psychometricians have their own set of ethical standards. They are uncomfortable with the fact that the data they worked so hard to generate are used as a tool to achieve political ends. From the OECD perspective, the more the PISA data results are used for policymaking decisions, the more visibility the OECD PISA receives, then the more relevant the OECD remains as an international institution and as a policy actor.

Modifying the Assessment

Changes were made to the original Data Strategy that was developed in 1997. These included the development of optional modules which would be "progressively introduced over multiple assessments"; emphasis on the "development of new assessment areas"; and the "opening up of possibilities for changing the target age for the younger age group" (OECD/EDU 2005c : 4). Several delegates view such changes as diluting the original overriding purpose of the OECD PISA, which was to measure the cumulative yield of educational
systems over time and to be a regular data source for the OECD educational indicators on student outcomes (Interview Respondent #6, 29 August 29 2006).

Expanding the OECD PISA to include a younger cohort directly challenges a territory in the international assessment field that has belonged to the IEA studies. In fact, the U.S. Department of Education strongly opposed expanding the OECD PISA to a younger cohort since these students were already being tested by the IEA studies (Interview Respondent #6, 29 August 2006). Because the U.S. Department of Education funds 25% of the OECD PISA's international costs, it could still exert some pressure on how the OECD PISA instrument was to be defined and implemented.

Conclusion

Since many users and agencies interface with the PISA, it becomes important for the OECD PISA to maintain its relevance and legitimacy as a producer of official educational knowledge. The PISA exists in relation to how its users interact with it. Users access the PISA to compile educational statistics, conduct secondary research, report rankings, use data as performance measures, and guide decision-making. International organizations such as the World Bank use the PISA data to monitor progress of educational project funding. These PISA users rely on the PISA remaining a reliable, valid and legitimate student assessment.
In this chapter, I pointed to the globalized marketplace that quantifies and standardizes student knowledge. This marketplace exists as part of the global architecture of education, reinforcing its network of power relations.

I argued that controversies and debates that were inherent in national and international student assessments have contributed to the fragility of the PISA. I discussed a number of issues that arise out of doing categorization and codification work. I also have pointed to the weaknesses of international student assessments and to the PISA.
Introduction
The dissertation traced the construction of an international student assessment, examining the practice of international student assessment seen as one aspect of the global architecture of education. The Programme for International Student Assessment (PISA) serves as a mechanism for legitimating, disseminating, and regulating a reconstructed form of official educational knowledge. Couched in OECD discourse, the PISA represents the role of education as the acquisition of competencies and skills in a knowledge based economy. Agents, acting from within the global architecture of education and armed with PISA data, can evaluate the outcomes of educational reforms and the efficiency of their own national educational systems.

In this chapter, I provide a summary of the dissertation’s findings and revisit the arguments that I made. I conclude with the implications of my findings and discuss the directions of future research work.

Dissertation Overview
In this section, I provide an overview of the dissertation and touch on some of the main ideas that were developed in earlier chapters. The dissertation described the OECD PISA as a power bloc formation that exercises power as it builds its technical capacity, relations of communication and power relations. Chapter 3 showed how the technical capacity for the science of educational
measurement was built by providing an overview of several developments in the quantification and measurement of 'intelligence.' It brought out the main debates and controversies surrounding the field of educational measurement and the theoretical foundations underpinning testing theories. I argued that these controversies, which would later undermine the stability of the practice of assessing students internationally, shed light on the fragility of such a practice.

Chapter 4 described how a science for international educational measurement was rendered doable and showed how American educational researchers influenced the direction of research in international comparative education settings. The creation of the International Association for the Evaluation of Educational Achievement (IEA) embodied the ideal for an empirically based science of comparative education that was influenced largely by the American model of educational research. The analysis brought out the controversies and debates that arose in the field of comparative educational measurement. They were never eliminated and in fact continue to contribute to the system's vulnerability.

Chapter 5 traced the development of OECD technical capacity for gathering educational statistics by looking at OECD educational activities under Keynesianism and neoliberalism. We saw how OECD educational activities were initially concerned with developing a set of tools for educational planning purposes as states invested in expanding their educational systems. During the neoliberal era, the OECD played an important role in the dissemination New Public Management ideas and in instilling these ideas in policy practices.
Educational policy directions turned away from investment in educational capacity to monitoring quality and effectiveness of educational systems. I showed how the OECD's technical capacity for educational measurement was also accompanied by the development of its relations of communication as the OECD built an infrastructure and coordinated activities in this area. The OECD's educational activities were integrated into regional and global educational networks.

Chapter 6 examined how American domestic politics played an influential role in shaping supranational educational directions under the Reagan (1981-1989) and George H. W. Bush (1989-1993) administrations. I argued that American involvement influenced the trajectory of the power bloc formation. The U.S. Department of Education was actively engaged in creating a comparative system of educational indicators and an international student assessment for measuring student achievement. The statistical measures reinforced the educational accountability systems that were being put in place in most OECD member states. When the Clinton Administration assumed office in 1993, it continued the previous administration's policies, albeit with an increased emphasis on the use of multilateral organizational structures to achieve American global interests. The U.S. Department of Education implemented its strategy for comparing student achievement internationally by using the IEA as its venue. Thus, in 1995, an international assessment for assessing students in mathematics and science was successfully administered by the IEA. The chapter also discussed the creation of an international adult literacy survey that
directly assessed the functional literacy of adults and provided a new way for quantifying the object ‘literacy.’

Chapter 7 focused my analysis on the conditions that led to the emergence of the PISA under the auspices of the OECD. I discussed how the PISA reflected a new problematization for internationally assessing students. The construction of the PISA was indicative of the pre-occupation of members of the ‘Rich Man’s Club’ in measuring the state of their human capital to ensure they had a competitive workforce for their knowledge-based economies. The PISA provided a set of indicators for measuring the quality of skills of a global workforce that was integrated increasingly into a global market economy. The chapter described the politics that underpinned the PISA’s creation and the rise of a new group of actors in the global architecture of education.

Chapter 8 described the relationships of power that constitute the bloc formation. I examined how the PISA instrument governs the conduct of individuals and schools by constructing the object ‘literacy.’ I argued that the PISA constructs the knowledge worker by enforcing a new curriculum template embodied in the abilities it tests. I compared the PISA testing approach to that of the IEA studies and concluded that the IEA student assessments govern the conduct of students at the national level since they are constructed to reflect national curricula. I analyzed the institutional power relations at the OECD and the IEA. I examined the OECD’s organizational structure by looking at how it functions as policy actor, policy forum and policy instrument. I argued that tensions arise out of these conflicting functions. I examined the IEA's
organizational structure and noted that it has strengthened its technical capacity over time and built its relationships within the global architecture of education. I argued that the IEA has a strong hold on disciplinary knowledge in the field of international student assessment which gives it a strategic advantage over the OECD.

In Chapter 9, I returned to two important themes that helped structure my analysis. I focused on how the PISA interfaces within a global network of users. I examined how the PISA is a node within the global architecture of education. The chapter identified the development of a profit-seeking market-driven educational measurement industry that operated on a world-wide basis. The chapter also identified the weaknesses and instability that characterized the field of international student assessment. In part, the PISA's fragility stems from the ongoing debates among experts as to the "best way" to test student achievement. In addition, the PISA's fragility is exacerbated by its location within a highly politicized environment. In particular, its status as a decentralized project within the OECD, made it dependent on contributions from participating states. This meant that the existence of the PISA could be jeopardized should the U.S. Department of Education withdraw from the program. The US remains the largest contributor to the OECD, but under the George W. H. Bush administration (2001 – present), the U.S. Department of Education began to reduce its level of involvement in the international comparative indicators project. These actions reflected the shift in American priorities away from international educational indicators to domestic concerns for monitoring.
educational quality as part of the No Child Left Behind legislation. For the latter, the IEA studies remained more relevant to the needs of American policymakers both federally and at the state level since they had similar test items as the U.S. National Assessment for Educational Progress.

The Dissertation's Arguments

The thesis advanced two key claims. First, the practice of testing students internationally forms an integral component of the global architecture of education. Second, programs such as the PISA nonetheless remain fragile and susceptible to contestation. Although educational measurement practices constitute techniques for reinforcing the official knowledge that certain powerful groups in society espoused, these hegemonic practices are nevertheless always unstable and fragile. Even though the results from student assessments are used to justify educational reforms and criteria for a new form of official knowledge, the indicators that were being used were socially constructed objects. They were created through a negotiated process by an elite group of experts and policymakers.

I supported these claims by examining the history of the construction of educational measurement. By delving into the history of student and intelligence testing and discussing how these techniques were rendered doable in the national context, I showed that the science of educational measurement was a societal construction. In the 1960s, national testing procedures were transferred to the international context and adapted to the practice of testing students.
internationally. National work practices and tools were adjusted so that they could incorporate a larger, cross-cultural sample of student populations. I highlighted the controversies and debates associated with testing students nationally and internationally and to the competing theoretical claims that were made by practitioners in this field.

The theme of American influence in the international arena was brought out through an examination of American domestic educational developments under the Reagan and George H. W. Bush administrations. Educational policies that focused on standards, accountability, quality and excellence were being adopted by OECD member states. The shift in educational policy directions required a new set of international indicators that focused less on educational inputs and more on educational outcomes. The U.S. Department of Education initiated the creation of an international educational indicators project under the auspices of the OECD. These educational indicators represented some of the data that fed into the global architecture of education along with other indicators produced by UNESCO and the International Association for the Evaluation of Educational Achievement (IEA).

With several of the OECD member states experiencing economic growth in high technology and information management sectors, there was a shift in the official knowledge discourse that informed the purpose of education. OECD member states had addressed the relationship between school and work in their vocational educational programs. In the 1990s, however, the focus was expanded to all students requiring a set of skills for the knowledge-based...
economy. In effect, the whole school system was being 'vocationalized' to ensure students were ready for the world of work. The creation of the Programme for International Student Assessment (PISA) reflected the OECD member states' preoccupation with measuring these types of skills. The dilemma that OECD member states faced was how to effect change in their own educational systems so as to render their own workers more competitive? The PISA was seen as a solution to this dilemma. It was to provide the 'objective' and potent data for legitimizing educational reform and for silencing opposition to such reforms, especially from teachers' unions.

In order to support my claim that the PISA formed an integral component of the global architecture of education, I showed how it reconstructed official knowledge and how it connected with the global educational architecture. The PISA's effect on the global architecture of education was to align closely student learning to the skills needed for economic growth in a knowledge-based economy. The PISA reinforced the human capital rationale for education. Even though the PISA was not a curriculum based assessment, its conceptual framework and test items delineated core skills and competencies that national curricula should be teaching. Experts involved in the construction of the PISA attempted to create conceptual frameworks and test items that would eventually set the standards for what students should be taught at school. The official educational knowledge that was to inform future educational reforms was being determined by a select group of individuals and agents that were connected to the global architecture of education. Even though the OECD PISA discourse
described the process as 'collaborative' among states, the construction of such international assessments was a technocratic process lacking in democratic input, particularly from parents, teachers, and students.

The PISA, which had inherited the practices, tools and techniques of earlier student assessments, nonetheless remained a fragile entity. The PISA was conceived using a technocratic framework that was devoid of the lived and complex realities of students, teachers and parents. I brought out the technical difficulties, controversies and debates that arose during the PISA's construction and how they were (temporarily) resolved by expert groups. The PISA, moreover, was constructed through a series of negotiated political compromises among representatives from OECD member states. I pointed to how the PISA was even more susceptible to contestation given the politicized environment it was located in.

The Implications of my Research

By focusing on the role of the OECD in the construction of the Programme for International Student Assessment, I shed light on how discursive and material practices were created and circulated within the global architecture of education. I was able to use Foucault's conceptualization of the power bloc formation to help structure my dissertation and my analysis. The political economy literature helped me locate the PISA within the broader political economic structure of the neoliberal project. Sociological concepts of power and science and technology helped me describe and analyze how an entity such as the PISA emerged, how
it was created, how it grew to become an intrinsic component of the global architecture of education and how it was destined to be a fragile construction.

By integrating the sociological concepts into public policy research, I was able to describe in concrete ways how something abstract, such as official knowledge, was constructed and disseminated through a power bloc formation of technical capacity, relations of communication and power relations. It was important to situate the PISA within a critical analysis of the global architecture of education in order to reveal the global power relations in which the PISA was implicated. By locating the PISA as an integral component of a larger dominant structure, I was able to trace in a more systematic manner how boundary objects, standardized packages and immutable mobiles were vital to the operation and sustenance of a system that manufactured official knowledge globally and circulated this knowledge nationally.

Once the power bloc formation of the PISA was traced, it was easier to discern its weaknesses. The OECD and the other international organizations implicated in the global architecture of education no longer appeared as formidable institutions. The research findings show that the PISA remains a fragile object – a result of a regulated adjustment within a power bloc formation. This suggests that such a formation can be undone when another adjustment takes place.
Future Research Directions

If we wish to reclaim a locally informed and contextualized understanding of education, then we need to trace, monitor and document the effects of the standardizing practices of entities such as the PISA. The official knowledge that is promulgated at the supranational level is transmitted to other levels of governance. Issues relating to education are analyzed through the prism of individual performance and achievement rather than the context of the lives of students, parents and teachers, the communities they live in and the inequalities that structure their lives. In the case of the PISA, we need to find out and document how this form of informational infrastructure operates locally and how it reinforces the official knowledge of OECD member states and ignores localized knowledge practices. As I have demonstrated, the PISA is a fragile construction that can be contested and destabilized through concerted oppositional efforts.
## Appendix 1

### List of Interviews

<table>
<thead>
<tr>
<th>Interview Respondent</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>October 12, 2005</td>
</tr>
<tr>
<td>#2</td>
<td>October 21, 2005</td>
</tr>
<tr>
<td>#3</td>
<td>November 24, 2005</td>
</tr>
<tr>
<td>#4</td>
<td>November 25, 2005</td>
</tr>
<tr>
<td>#5</td>
<td>May 2, 2006</td>
</tr>
<tr>
<td>#6</td>
<td>June 5, 2006 &amp; August 29, 2006</td>
</tr>
<tr>
<td>#7</td>
<td>June 9, 2006</td>
</tr>
<tr>
<td>#8</td>
<td>June 19, 2006</td>
</tr>
<tr>
<td>#9</td>
<td>June 29, 2006</td>
</tr>
<tr>
<td>#10</td>
<td>July 8, 2006</td>
</tr>
<tr>
<td>#11</td>
<td>July 10, 2006</td>
</tr>
<tr>
<td>#12</td>
<td>August 15, 2006</td>
</tr>
<tr>
<td>#13</td>
<td>August 24, 2006</td>
</tr>
<tr>
<td>#14</td>
<td>August 27, 2006</td>
</tr>
<tr>
<td>#15</td>
<td>August 29, 2006</td>
</tr>
<tr>
<td>#16</td>
<td>August 30, 2006</td>
</tr>
<tr>
<td>#17</td>
<td>September 18, 2006</td>
</tr>
<tr>
<td>#18</td>
<td>September 22, 2006</td>
</tr>
<tr>
<td>#19</td>
<td>September 24, 2006</td>
</tr>
<tr>
<td>#20</td>
<td>October 20, 2006</td>
</tr>
<tr>
<td>#21</td>
<td>October 25, 2006</td>
</tr>
<tr>
<td>#22</td>
<td>October 31, 2006</td>
</tr>
<tr>
<td>#23</td>
<td>November 10, 2006</td>
</tr>
</tbody>
</table>
Appendix 2
Consent Form

Informed Consent Form
Date of Ethics Approval: September 22, 2005
Title: The OECD PISA: Unraveling a Knowledge Network

I understand that I am being invited to participate in a research study conducted by Clara Morgan, candidate in the PhD program in Public Policy and Administration at Carleton University. I have read the letter of information that has been provided by Clara Morgan, and I understand that the information I provide will be used to inform and shape the researcher’s dissertation.

I understand that the purpose of this research study is to analyze and investigate the Organization for Economic Development and Co-operation’s (OECD) Programme for International Student Assessment (PISA). The study seeks to understand the conditions and events that led to the OECD PISA’s emergence, the ways in which the OECD PISA and the PISA results have been coordinated, administered, implemented and disseminated across multiple levels of governance. Of particular interest to the study is the politics and sociology of international large scale assessments and their implication on educational research and policymaking.

I understand that I will be providing information through an interview where I will be asked to discuss and describe my involvement in international large scale assessments, including the PISA. I realize that the interview may take up an hour of my time to complete and that I may decline to answer any questions. I understand that I am being asked to verify my permission to participate in this study with the administrator of the department I represent, if this is applicable. I understand that I may withdraw agreement to participate at any time during the interview or for up to 7 days after completion of the interview. At that time, I know that I may indicate whether or not the data collected up to that point can be used in the study, and that any information I do not want used will be destroyed immediately. After 7 days, I realize that the researcher may begin to use my responses to direct further research. I understand there is a possibility that the researcher may contact me for further clarification of my responses for a period of up to three months after the completion of the interview.

I understand that I have a choice of responding to the interview by telephone, email, fax, mail or in-person (if feasible). I understand that I am being offered confidentiality so that none of my comments, opinions, or responses will be attributed to me. I understand that the researcher will not include identifying information in any report, summary or presentation. I also understand that the interviews are being drawn from a small sample and even though my opinions and responses will not be attributed to me,
complete confidentiality cannot be ensured. This is the only identifiable risk that is associated with the study.

I understand that this interview will be audio recorded in order to ensure accuracy of the information collected. I also understand that I can choose not to consent to be recorded and that the researcher will then take interview notes. I realize that this study may not directly benefit me or the organization I work for.

I understand that the data collected will be seen only by the researcher, and that my responses will be stored in a locked filing cabinet until a year after the researcher’s dissertation defense. At this time, I understand that the data will be destroyed using a mechanical shredder.

I understand that research findings will be presented by the researcher in conference papers, published journal articles, published chapters in books and the dissertation thesis.

I understand that this dissertation research study has been reviewed and received ethics approval from the Carleton University Ethics Committee and from other relevant research advisory committees. If I have questions or concerns, I know that I may contact the Carleton University Research Ethics Committee chair:

Professor Antonio Gualtieri 613.520.2517
ethics@carleton.ca

Questions and concerns may also be directed to the following individuals:
Clara Morgan, Researcher 613.823.0749
claracmorgan@rogers.com

Dr. Bruce Curtis, PhD co-supervisor 613.520.2600 x2596
bruce_curtis@carleton.ca

Dr. Rianne Mahon, PhD co-supervisor 613.520-7414
rmahon@ccs.carleton.ca
(on sabbatical until July 2006)

Do you consent for the interview to be audio recorded?
Yes ___  No ___

Do you agree to participate in the study?
Yes ___  No ___

Name of participant: ____________________________

Signature of participant: ____________________________

Date: ____________________________
Appendix 3
List of Questions

The interview respondents were provided with the following list of questions to help guide the interview.

- How did the idea of an OECD assessment programme emerge? When did this idea first become articulated?

- Describe your involvement in the PISA.

- What kinds of challenges and successes have you faced during your involvement with the creation of the PISA?

- How have you dealt with these challenges and successes, either individually or with others?

- Which areas of contention arose during your involvement with the PISA and how were they resolved?

- What is your opinion of the PISA as compared with other international assessments?

- What do you deem to be the effects of the PISA, both positive and negative?

- Is there anything else you would like to discuss with me in relation to the PISA that you find relevant to my research?

- Are there any specific publications and data sources you would like to recommend to me?

- Are there certain individuals you recommend I should speak to?
References

Published Books, Reports and Articles


Binet, A. (1903) *L'étude expérimentale de l'intelligence.* http://gallica.bnf.fr/ark:/12148/bpt6k758553/f309.table


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Other References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Available at: http://cisad.adc.education.fr/revapdf/lire20.pdf


OECD (1997b) Data Strategy for the Development of Student Achievement Indicators on a Regular Basis. Terms of Reference for the Tendering Procedure. [DEELSA/ED/CERI/CD(97)6]


OECD/DEELSA (1997a) First meeting of the Board of Participating Countries, Paris, 6-7 October. [BPC(97.1).99bis]

OECD/DEELSA (1997b) Second meeting of the Board of Participating Countries (BPC), Paris, 15-16 December. [BPC(97.2).99]

OECD/DEELSA (1997c) Meeting of the Technical Expert panel for the revaluation of proposals to the OECD call for tender BPC(97.1)7. OECD: Paris. [BPC(97.2)2]
OECD/DEELSA (1998a) Third Meeting of the Board of Participating Countries (BPC), San Francisco, 20-21 April. [DEELSA/PISA/BPC(98)8]

OECD/DEELSA (1998b) Fourth meeting of the Board of Participating Countries (BPC), Paris, 6-7 July [DEELSA/PISA/BPC(98)16]

OECD/DEELSA (1998c) Fifth meeting of the Board of Participating Countries (BPC), Paris, 29-30 October [DEELSA/PISA/BPC(98)32]


Available at: http://www.oecd.org/dataoecd/17/17/35378941.pdf

Available at: http://www.oecd.org/dataoecd/56/34/35557211.pdf


OECD/INES (1997c) NWA-BPC Transition Meeting Record. Transition Meeting – Toward the Board of Participating Countries. May 7-8.


OECD/INES (1998) Strategic Plan for Network A.


