



UNIVERSITY of OULU  
OULUN YLIOPISTO

MURPHY, DANIEL

THE GLOBAL EDUCATION REFORM MOVEMENT AND THE OECD'S  
PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT (PISA): A  
SECONDARY ANALYSIS OF PISA DATA

Master's Thesis in Education

FACULTY OF EDUCATION

Master's Degree Programme in Education and Globalisation

2012

## TABLE OF CONTENTS

1. INTRODUCTION .....	1
1.1 The global education reform movement .....	1
1.2. The Programme for International Student Assessment (PISA) .....	2
1.3 Purpose and structure of the research.....	2
2. BACKGROUND TO THE RESEARCH .....	4
2.1 Nature and extent of the global education reform movement.....	4
2.2 Arguments in favour of the global education reform movement.....	7
2.3 Criticisms of the global education reform movement.....	8
2.4 Influences and actors behind the spread of the global education reform movement .	10
2.4.1 Globalisation.....	10
2.4.2 Human capital concepts .....	11
2.4.3 New public management .....	12
2.4.4 Preference for scientific and quantitative modes of analysis .....	12
2.4.5 Public appeal.....	13
2.4.6 International organisations .....	13
2.4.6.1 The World Bank.....	14
2.4.6.2 The Organisation for Economic Cooperation and Development (OECD) .	15
2.4.7 Neoliberal ideology .....	16
2.5 Summary .....	18
3. DISCUSSION OF RESEARCH METHODS.....	20
3.1 Quantitative non-experimental research.....	20
3.2 Correlational research .....	21
3.3 Secondary analysis .....	22

3.4 Appropriateness of current study, methods used and research questions .....	24
4. PISA OPERATIONS AND DATA COLLECTION .....	26
4.1 Background .....	26
4.2 Administration.....	27
4.3 Translation and cultural appropriateness.....	29
4.4 Quality assurance .....	30
4.5 Sampling.....	31
4.6 Maintaining precision in the sample .....	32
4.7 The PISA test, coding and national mean scores .....	33
4.7.1 Science.....	34
4.7.2 Proficiency levels and question difficulty .....	37
4.7.3 Reading.....	40
4.7.4 Mathematics.....	42
4.7.5 Coding of Students' Answers .....	43
4.7.6 Weighting .....	44
4.8. Contextual Information .....	44
4.9. Discussion of PISA methodology .....	52
5. 2006.....	57
5.1 PISA's Analysis and Results.....	57
5.1.1 Science Performance and School and System Characteristics .....	58
5.1.2 School Competition .....	58
5.1.3 School accountability policies .....	59
5.1.4 School autonomy .....	60
5.2 Secondary analysis using PISA 2006 data .....	62
5.2.1 Development of research design.....	62
5.2.2 Independent variables .....	64
5.2.3 Dependent variables .....	65

5.2.4 Results of secondary analysis using 2006 data.....	65
5.3 Summary .....	67
6. 2009.....	68
6.1 PISA's Analysis and Results.....	68
6.2 Secondary analysis using PISA 2009 data .....	71
6.2.1 Independent Variables .....	71
6.2.2 Dependent variables .....	72
6.2.3 Results of secondary analysis using 2009 data.....	72
6.3 Summary .....	74
7. CONCLUSIONS.....	76
8 . REFERENCES .....	78
APPENDIX.....	86

## **1. INTRODUCTION**

In an era of globalisation, education policy formulation is subject to different influences. While national and local economic, political and cultural conditions are relevant, global trends are becoming more influential (Burbules and Torres 2000). Global convergence of education policies is a reality, and shaped by influential attitudes and actors that operate on an international scale. These attitudes and actors can be said to be part of a global education reform movement (GERM). The Programme for International Student Assessment (PISA), a triennial test of students' abilities in reading, mathematics and science, is one of the actors influencing education policy on a global scale. By subjecting PISA's work to secondary analysis this study addresses two research questions:

1. What does secondary analysis of PISA data reveal about the relationship of a range of typical GERM school organisation policies to student learning outcomes?
2. What does this analysis reveal about PISA and its relationship to the GERM?

### **1.1 The global education reform movement**

Sahlberg (2007) uses the term global education reform movement to label the set of school organisation policies that have become prevalent over the last 30 years in many countries. This has been a period of marked policy convergence with a number of commonly recurring features: standardised testing of students, schools and teachers made accountable for results of tests, schools given more autonomy and encouraged to compete, a drive for greater efficiency and promotion of private provision. Not all countries are affected and some implement some elements and not others, but there has been an undeniable trend across the globe in the direction of these policies. Supporters of these policies make a range of claims that they can make schools systems more efficient and responsive to students and their families, and that they enhance student learning. As will be seen, these claims are contested.

In a time of intensified globalisation, convergence of education policies – indeed any policies – is inevitable. The form this takes is shaped by dominant attitudes and actors of

the era. In this paper I use the term global education reform movement to refer to the policy agenda outlined above and those actors and attitudes that are behind it.

## **1.2. The Programme for International Student Assessment (PISA)**

The Programme for International Student Assessment is a triennial common test and data gathering exercise carried out in 65 countries (2009) by the Organisation for Economic Cooperation and Development (OECD). PISA's roles as reference for policy makers and source of stories for media outlets have given it great influence in education policy debates (Grek 2009, Figazzolo 2008). PISA publishes detailed descriptions of student achievement within and between countries in the tested subject domains of reading, mathematics and science. With the large amount of contextual data it gathers, PISA is able to investigate relationships between such matters as students' socio-economic status, student attitudes and school organisation policies, and learning outcomes as measured by its test. In matters of school organisation, PISA collects data on the degree to which countries implement a wide variety of policies including many which are part of the GERM. PISA's analysis of its 2006 data led to the key findings that having a larger number of schools that compete for students is associated with better results, that there is a significant positive association between schools making their achievement data public and having stronger results, and that students in countries where autonomy is more common tend to do better. PISA 2009 found that in countries where schools have greater autonomy over what is taught and how students are assessed, students tend to perform better.

## **1.3 Purpose and structure of the research**

As background to the study I establish the existence and nature of the global education reform movement. This movement is best understood as the net effect of a variety of attitudes and actors. In Part 2 I seek to provide a comprehensive account of the factors that contribute to the GERM. These are numerous and often act in complimentary and reinforcing fashion.

Issues surrounding the research method to be used – correlational statistics – are then discussed, as is the fact that the study constitutes secondary analysis of PISA. Secondary

analysis has a well established position within educational research and is becoming more widespread as PISA expands in size and influence.

When undertaking secondary analysis it is crucial that the researcher has a complete understanding of the original study. Part 4 comprises a detailed description of PISA's background, administration, field operations, data collection and how it quantifies its variables. The section concludes with a review of literature discussing PISA's methodology.

Before presenting the results of my research I look at how PISA approached its analysis of relationships between school organisation policies and student performance for the 2006 survey and how they presented their findings. I repeat this for 2009, noting that PISA changed its methods of statistical analysis. In fact, it adopted the widespread use of cross-country correlations that I first used in my analysis of the 2006 data.

My results raise questions about the merit of consequential accountability policies favoured by the GERM. These results, coupled with a detailed analysis of how PISA chooses to use the data it collects and treat its findings, raise important questions about the objectivity and credibility of PISA.

## 2. BACKGROUND TO THE RESEARCH

### 2.1 Nature and extent of the global education reform movement

Sahlberg (2007) uses the term global education reform movement (GERM) to refer to school policies that have increasingly been adopted in many parts of the world since the 1980s with the aim of improving the quality of education.

In *Education Policies for Raising Student Learning: The Finnish Approach* (2007) Sahlberg describes this agenda as having at least three common features:

- 1) The standardisation of education: The setting of performance standards and testing to evaluate their attainment through frequent external assessment measures.
- 2) Increased focus on literacy and numeracy: Basic knowledge and skills are considered most important and are also more suited to standard testing and comparison.
- 3) Consequential accountability: The tying of school performance to the processes of accrediting, promoting, inspecting and rewarding or punishing schools and teachers.

A more comprehensive description of the policies around which education is converging globally would also include Sahlberg's characterisation of the essence of globalisation of education in *Teaching and Globalization* (2004): decentralisation, privatisation and the drive for increased efficiency in school systems.

The global convergence in schools policy on the terms described by Sahlberg has been widely noted. Forsey (2007) writes of a new 'global education policy consensus' in favour of school autonomy linked to efficiency and cost saving and the focus on pre-specified outcomes and measurable competencies. Robertson (2008) describes changes to education systems as a remandating which forces teachers and schools to demonstrate their effectiveness through national and global testing systems, use funds more efficiently, increasingly seek private sources of income and accept performance based 'merit' pay systems. Steiner-Khamsi (2004) notes how of all the possible school reform models, a select few – school choice and competition, outcomes-based education and standards – continue to surface in different parts of the globe. Outcomes-based education became popular in the 1980s and was supplanted in the 1990s, according to Sahlberg

(2004), by standards based education. Together these trends shifted the focus from educational inputs such as regulation of teaching to the monitoring of learning.

Carnoy and Rhoten (2002) write of an ideological package consisting of decentralisation and privatisation, choice and accountability and testing and assessment having an impact on education systems worldwide. Hargreaves (2003) and Ball (1998) refer to a new orthodoxy of education to describe this convergence. Rinne, Kivirauma & Simola (2002) observe a new type of consensus among the political elite in favour of decentralization, goal steering, accountability, managerialism, evaluation, choice, competition and privatisation (p.643). Moutsios (2009), Green (2006), Grubb and Lazerson (2006), Hargreaves (2003, 2006) Ravitch (2010) Ball (2006) and Kamens and McNeely (2010) are others to have described the spread of education policy with the common characteristics described above.

A 2007 UNESCO survey found national assessment regimes were spreading across all regions of the world. Between 1995 and 2006 the number of countries with standardised national testing regimes more than doubled from 28 to 57, with the trend greatest in the developed nations of North America and Western Europe (Benavot & Tanner 2007). The manner in and degree to which GERM policies are implemented in a national context are determined by local factors. These include a country's economic strength and reliance on international lending agencies, which often attach education reform conditionality to aid, and local politics and traditions including the presence of protagonists in favour of such reforms, and resistance to them (Ball, 1998, Sahlberg 2004, Rinne, Kivirauma & Simola 2002). The fullest embrace of GERM policies, involving well developed education markets and strong consequential accountability systems, has taken place in England, New Zealand, the United States, Chile, France, Colombia and parts of Australia and Canada. Portugal and Sweden are also proceeding in this direction (Ball 1998, Sahlberg, 2004). A larger and growing number of countries are adopting elements of the template.

In the United States the federal No Child Left Behind Act of 2001 required all schools in the country that receive federal funding to make adequate yearly progress on state-run standard tests with the aim of having every student in the country proficient in reading and mathematics by 2014. Failure to make adequate yearly progress triggers interventions including requiring the school to facilitate parents' choice to take their children to a

school making the required progress, replacing all or most staff or handing control of public schools to private operators (Hamilton, Stecher, Vernez & Zimmer 2007).

The spread of the GERM template among nations has been accompanied by a growth in international testing and benchmarking. This began with the International Association for Evaluation in Education's Trends in Mathematics Study in the early 1990s. In 1992 the Organisation for Economic Cooperation and Development (OECD) published its first volume of *Education at a Glance: Education Indicators*. This collated demographic, resource and outcome data collected by OECD member countries. The Indicators in Education (INES) project continues in expanded form alongside the OECD's tri-annual Programme for International Student Assessment (PISA). 43 countries took part in the first PISA survey in 2000. By the time of the latest survey in 2009 that number had risen to 65, comprising 34 OECD members and 31 partner countries. Student achievement data for mathematics, reading and science, and contextual information are used to rank countries' performance and investigate relationships between students' backgrounds and education policies and learning outcomes (OECD 2010b).

PISA's triennial basis allows for the monitoring of education policy trends over time. For the period between the 2006 and 2009 surveys more than half of the participating countries reported a reduction in restrictions on school choice. Twelve OECD countries reported the creation of new more autonomous school models and 10 had introduced new funding mechanisms to promote school choice and competition (OECD 2010a).

The existence of pervasive common global education reforms is thus well established and well documented. The question of whether this international trend comprises a movement is an interesting one. Sahlberg (2007) uses the term movement to describe a common set of policies, or policy objectives. However as he and many others describe, and as is discussed below, the spread of these policies is explained by an interrelated web of actors and factors. Some sociological definitions of social movements centre on the actions of multiple actors, be they individual or institutional, to achieve change (Hoult 1969, Diani and Della Porta 1992). A broader definition sees social movements as 'a set of attitudes and self conscious actions by people who seek to change society's structure or ideology' (Conklin, 1984 p. 460). When using the term global education reform movement in this study I refer to the combination of beliefs and actors that are driving the convergence of

schooling around what I call GERM policies. The GERM policies can be best summarised as

- Emphasis on teaching the ‘basic’ subjects, literacy and mathematics
- Regular standardised testing of these subjects
- Results of this testing made public to enable comparison
- Schools given more autonomy over curriculum and management but held accountable to achieve mandated objectives through test results and administrative targets
- Promotion of private schooling and the idea that schools need to become more efficient and seek private sources of income
- Belief that market forces and competition will improve schooling and are the best way to achieve efficiency
- Promotion of merit/performance pay for teachers, and antipathy to teacher unionism and collective bargaining.

Not all countries are implementing this agenda and some are implementing parts of it. Nevertheless there has been indisputable convergence around these policies in the last 30 years.

## **2.2 Arguments in favour of the global education reform movement**

Arguments in favour of GERM policies are usually advanced in terms of improving educational quality and student outcomes, and making schools more efficient. Sahlberg (2004) writes

Standardization-oriented reforms...are based on the assumption that in competitive economic and social contexts the quality of education...can best be improved by setting high performance standards for teaching and learning then measuring whether these standards have been met (p.72)

Publishing of test results and other indicators places pressure on schools and teachers to improve test outcomes. Jacob (2005), in a study of test scores in Chicago after an accountability policy was implemented in 1996, found maths and reading achievement as measured by standard tests increased sharply.

Making test information public also enables parents to make an informed choice in the context of a competitive schools market (Apple 2004). In the early 1990s in the United States, Chubb and Moe's influential *Politics, Markets and America's Schools* (1990) urged policy makers to see choice as a panacea that would make schools more accountable. Competition, it was argued, improves schools' efficiency and quality, as it does in the business sphere. Analysing Sweden's experience with increased school choice, Sandström and Bergström (2002) found school results improved as a result. In Britain, Bradley and Taylor (2002) found strong evidence that the quasi-market in schooling that developed there in the 1990s led to a substantial improvement in efficiency.

Decentralisation is closely related to choice. Greater autonomy for schools is advanced in terms of making them more responsive to local needs but it is also a prerequisite for a schools market (Sahlberg 2004, Ravitch 2010).

Arguments and findings in favour of core GERM policies are summarised in Hanushek and Woßmann's 2008 paper for the World Bank, *Education Quality and Economic Growth*.

- Consequential accountability including merit pay for teachers will improve student performance.
- Countries with a larger share of private schools tend to perform better.
- Choice and competition puts pressure on teachers and schools to perform lifting student achievement. Their introduction in the Czech Republic and United States has had positive effects on student performance.
- School autonomy is essential to establish an incentive system and has been shown to improve student learning outcomes.
- Strong accountability systems must accompany choice and autonomy and can lead to better student performance.

### **2.3 Criticisms of the global education reform movement**

The spread of GERM policies has not been without controversy. Common criticisms of standard testing and consequential accountability systems are that placing such importance on test results leads to teachers and schools teaching to the test and narrowing

their curriculum around those areas that are tested. Rather than increased diversity and responsiveness to local needs, what results is greater uniformity. Also neglected are education's broader goals (Sahlberg 2004).

Ravitch (2010) writes that in the United States, claims by advocates of consequential accountability that it leads to improvements in learning do not stand up to scrutiny. What has emerged however is 'gaming' of results through excluding struggling students from tests, selective admission procedures and cheating. The plateauing of results several years after the introduction of national tests in England is, according to Thrupp and Hursh (2006), evidence that teachers adjusted to teach to the test but this approach delivered diminishing returns. They go on to state that 'reforms such as target setting divert attention from more fundamental solutions, they are themselves deeply damaging' (p.653).

Sahlberg (2004) writes there is no evidence that school autonomy increases student achievement and experiences from New Zealand, Chile and the United States where privatisation and competition have been introduced do not support claims that they improve learning. Chile's experience was also examined by Carnoy and McEwan, (2000) who found promotion of a schools market did little or nothing to improve academic performance.

Marsh, Springer, McCaffrey, Yuan, Epstein, Koppich, Kalra, Di Martino and Peng's (2011) survey of New York's Schoolwide Performance Bonus Program found no evidence of improved student performance related to the tying of school bonuses and penalties to students' test results. Amrein and Berliner's (2002) analysis of national test results in 18 states in the USA found student learning remains at the same level or goes backwards when high stakes testing is introduced.

Analysing experiences in New Zealand and England, Apple (2004) found that school choice and accountability through standard testing are likely to exacerbate differences in access and outcomes based on race, ethnicity and class. He believes that schools rather than families end up exercising choice. That is they admit students they believe will enhance their test scores and prestige at the expense of ethnic minorities and students with special needs.

Ball (2006) writes that constant appraisal of their work and tying of salaries to student test results have been criticised for creating demoralisation, insecurity and guilt among teachers, and making the profession less attractive.

## **2.4 Influences and actors behind the spread of the global education reform movement**

The global education reform movement is best understood as the net effect of a number of concurrent and in many cases reinforcing influences. Some are influential in certain countries and not in others, and some are global phenomena.

### **2.4.1 Globalisation**

Internationalisation of capital, trade, industrial organisation, communications and political institutions since the middle of the twentieth century have increased co-ordination and interdependence of many important activities on a world scale (Tilly, 2004). This globalisation has involved not only greater integration of economic activity but also of many other areas of social organisation. Meyer, Boli, Thomas and Ramirez (1997) write of the emergence of a 'world society' after World War 2 with nations increasingly drawing on international frameworks for national policies. Policymaking is increasingly globalised and education is no exception (Moutsios 2009). Advances in communications and increased student and staff mobility encourage policy borrowing (Green 2006).

Convergence under these circumstances is inevitable, but globalisation in a predominantly capitalist world favours certain policies over others. Monkman and Baird (2002) write of an 'undeniable interweaving' between globalisation and neoliberal economic ideas. The small state, low taxes and free market ideas favoured by neoliberalism can often be seen to go 'hand in hand' with globalisation effects (p.502). As is discussed in more detail below, neoliberal ideology that has become very influential in the era of globalisation has well developed ideas on education. Also, education is widely seen as having a key role to play in determining a nation's ability to compete in the globalised system (Green 2006, Grubb & Lazerson 2006, Kamens & McNeely 2010, Carnoy and Rhoten 2002).

## 2.4.2 Human capital concepts

The concept of human capital refers to the fact that human beings invest in themselves by means of education, training, or other activities, which raises their future income by increasing their lifetime earnings (Woodhall, 1987 p.219)

After being first articulated in the early 1960s this concept became of great use to education economists and planners as it allowed them to carry out cost-benefit analysis of education funding and calculate returns on investment, both private and social (Woodhall 1987). Also in the early 1960s, the United Nations General Assembly adopted a resolution stressing the importance of education for economic development (Resnik 2006). Today the main multilateral organisations setting global education policies – The World Bank, OECD, UNESCO - subscribe to a human capital view, meaning they see education not only as a measurable investment but that its main purpose should be to generate human resources that will provide an economic return to the individual and the nation (Moutsios, 2009). The World Bank has what Jones (2004) terms an ‘overtly ideological commitment to rate of return analyses...grounded in human capital theory’ which leads it to neglect aspects of education that do not provide a measurable economic return. This outlook encourages private provision of education to get costs off the balance sheets of governments with limited fiscal resources. (Robertson, Novelli, Dale, Tikly, Dachi and Alphonse 2007). According to Rizvi and Lingard (2006), the OECD is a key promoter of the concept of the knowledge economy; that knowledge is the ‘central intrinsic component of economic production and activity’ and accordingly the economic goals of education are given priority over its social and cultural processes (p.252).

What Resnik (2006) calls the ‘Education-Economic Growth Black-Box’ has gained great currency around the world, although it is often conveyed in the negative as a prelude to reform. That is, poor economic performance is blamed on ‘failing’ schools. This argument has often paved the way for the introduction of GERM-type policies (Grubb and Lazerson 2006). This was particularly the case after the economic crisis affecting developed economies in the 1970s (Forsey 2007), notably in the United States where the official report *A Nation at Risk*, published in 1983, blamed many economic and social problems on the education system and this is seen as pivotal in paving the way for GERM policies (Ravitch 2010, Hursh 2005).

The human capital/education-economic growth paradigm favours policies including concentrating teaching on the basics of literacy, mathematics and science, standardised measurement of education inputs and outcomes, private provision and the drive for efficiency.

### **2.4.3 New public management**

Over the last 30 years education has been increasingly influenced by ideas drawn from business management. This reflects a widely held view that business models of management are more efficient than bureaucratic ones, which is in turn part of broader public scepticism about the ability of government to look after people's needs (Forsey 2007). Devolution and quality assurance both found expression in corporate culture before being incorporated into school governance in many countries (Sahlberg 2004, Forsey 2007, Kamens and McNeely 2009). Preference for business-type models of public sector organisation reflects disillusionment with traditional state provision but also an ideological push, discussed in more detail below. In many countries, public sector management and service provision have been restructured under the guise of what is called 'new public management'. This entails a push for greater efficiency, empowerment of management at the expense of traditional bureaucratic procedures and 'steering at a distance', ensuring compliance with policy and efficiency goals through targets and benchmarks. The OECD has been a strong advocate of new public management in all areas of public provision (Lingard and Grek 2007, Rizvi and Lingard 2006).

### **2.4.4 Preference for scientific and quantitative modes of analysis**

The shift toward testing and measuring in education can be seen as part of a broader preference that technical and scientific methods guide policy. This 'hegemony of science' as Kamens and McNeely (2009) term it, demands so called evidence based decision making, something that favours quantitative data gathering and analysis over other forms. The United States' *No Child Left Behind Act* (NCLB), probably the fullest legislative expression of GERM policies anywhere in the world, contains more than 100 references to scientifically-based research (Lauer 2006). This preference feeds into what

Martens (2007) calls the comparative turn in education policy formulation. That is, analysis using quantitative indicators invites comparison, from which models of best practice are identified and transplanted into other contexts in a form of ‘government by comparison’ (p.54).

#### **2.4.5 Public appeal**

Greater school choice and autonomy are often framed as responses to widely held parent desires for more involvement in their children’s education. These proposals and the provision of parents with more information about how their child’s school performs have proven politically popular (Ravitch 2010). Certainly GERM advocates often put forward these policies in terms of empowering parents and responding to their demand for choice. (OECD 2007b, Robertson, Novelli, Dale, Tikly, Dachi and Alphonse 2007) According to Kamens and McNeely (2009) school devolution responds to wishes for more authority among professionals, non-government organisations and individuals. Egalitarianism and democracy are common ideals and elements of the GERM, such as local autonomy, choice and accountability are often framed in these terms by advocates. Implementing GERM-type reforms can thus be attractive to politicians who find their policy scope limited in other areas, like employment policy, by fiscal and political constraints (Green 2006).

#### **2.4.6 International organisations**

A thickening web of international organisations (IOs) including the World Bank (WB), International Monetary Fund (IMF), the World Trade Organisation (WTO) and the OECD increasingly define the process of globalisation (Robertson, Novelli, Dale, Tikly, Dachi and Alphonse 2007, Lingard and Grek 2007, Moutsios 2009). Developing countries have always found themselves subject to external policy influence through the demands of aid donors and lending agencies, but increasingly, developed countries are ceding policy autonomy to IOs (Rizvi and Lingard 2006). For Moutsios (2009) transnational institutions are now the main contexts which define the major educational aims for most countries. IOs, particularly the WB and OECD, driven in part by what Resnik (2006) calls an ‘intrinsic tendency to increase their power and resources’ (p.178) have taken a central

role in debates over education policy and are arguably the most influential actors in the growth of the global education reform movement (Kamens and McNeely 2009). What both Grek (2009) and Steiner-Khamsi (2004) call externalisation – the setting of national policy in reference to international examples or norms propagated by IOs - is now central to education and it is necessary to look at the policy objectives and methods of the two most influential actors.

#### **2.4.6.1 The World Bank**

Established as part of the Bretton Woods agreement following World War 2, the World Bank's (WB) charter has been to assist development and poverty alleviation through loans. Earlier, this took the form of financing large infrastructure projects, but influenced by the concept of human capital and the link between education and economic growth, the bank took a greater interest in funding education from the 1960s. (Heyneman 2003)

World Bank loans have always come with conditions attached, and this is the source of the bank's influence over policy in developing countries. In the 1980s, the bank's structural adjustment loans required recipient countries to implement a policy prescription of deregulation, competitiveness and privatisation which became known as the Washington consensus (Robertson 2008). Education was by this era a bank priority, and the bank insisted education be reformed on these lines, adding by the late 1980s, national testing and assessment regimes (Robertson, Novelli, Dale, Tikly, Dachi and Alphonse 2007). The bank's 1990 report *Improving Primary Education in Developing Countries* was a strong argument in favour of testing and monitoring (Kamens and McNeely 2009). With great zeal, the bank insisted its preferred education approach be implemented in all countries it dealt with, regardless of local conditions and traditions. It was influential in much of the world's poorest countries in this era, as well as in Eastern Europe following the collapse of communism (Jones 2004). Today the bank is the single biggest funding provider for education programmes (Heyneman 2003).

Slow economic growth and increasing social stratification led to disillusionment with the Washington consensus in the mid to late 1990s. The bank responded with what it called more of a social focus and emphasis on 'good governance' but its fundamental economic policy focus – market liberalisation, privatisation, competition – remains very similar

(Robertson 2008, Jones 2004). This continuity is reflected in education. The bank's 2008 Report *Education Quality and Economic Growth* restated the importance of the former to the latter, and called for choice and competition, school autonomy, consequential accountability for schools and performance pay for teachers (Hanushek and Woßmann 2008).

#### **2.4.6.2 The Organisation for Economic Cooperation and Development (OECD)**

Sometimes referred to as the rich man's club, the OECD's membership currently comprises 34 developed countries that meet the entrance requirements of running a market economy governed by a pluralistic democracy (Lingard and Grek 2007). The OECD exerts great influence over its members and indeed non-members, many of which aspire to developed country status and membership of the organisation. In Porter and Webb's (2007) words, the OECD 'defines the standards of appropriate behaviour for states which seek to identify themselves as modern, liberal, market-friendly and efficient' (p.3). The OECD's role as a research and advisory body is funded by members although contributions are uneven. Seven countries (the G8 minus Russia) contribute 80 per cent of funding with the United States providing 25 per cent and Japan 23 per cent (Moutsios 2009).

In its earlier years, the OECD debated different economic theories but after the 1970s economic crisis, it became dominated by a pro-free market or neo-liberal ideology, spelt out in its 1975 McCracken Report (Rizvi and Lingard 2006, Porter and Webb 2007). It has taken an increasing interest in education since the 1980s, and has done so from a strong human capital standpoint. In its own words, 'the prosperity of countries now derives to a large extent from their human capital and, to succeed in a rapidly changing world, individuals need to advance their knowledge and competencies through their lives' (OECD 2007b p.3).

In the 1980s, a major conflict took place within the OECD over the nature of its education work. The British and United States governments applied great pressure to make basic skills and the collecting and comparison of data on inputs and outcomes the focus of education policy research and analysis. While this push met resistance from OECD staff, it soon prevailed and the first volume of *Education at a Glance: Education Indicators*

was published in 1992 (Papadopoulos 1994). In the 21<sup>st</sup> century, the establishment of the Programme for International Student Assessment has developed this work but with data collected by the OECD itself. The apparent meticulousness and scope of PISA's work give it credibility and clout (Steiner-Khamsi 2004, Lingard and Grek 2007). It benefits from being perceived as autonomous, and above local politics, and in the words of Grek (2009) has 'achieved a brand that most regard as indisputable' (p.25).

The OECD is however not just a research organisation publishing disinterested reports. It is also a powerful policy actor advancing a strong human capital concept of education and asserting that consensus exists on many aspects of education policy required to respond to imperatives of globalisation. In voluminous reports, it mixes dry analysis with normative writing that contains many assumptions, and seeks to steer debate not merely inform it (Robertson, Novelli, Dale, Tikly, Dachi and Alphonse 2007, Rizvi and Lingard 2006, Kamens and McNeely 2009). For example, on the issue of school choice

When parents and students can choose schools based on academic criteria, schools then compete for students, which in turn, may prompt schools to organise programmes and teaching to better respond to diverse student requirements and interests and so reduce the costs of failure and mismatches (OECD 2010b p.72).

The measures PISA chooses to collect data on, and conduct and publish analysis of, also have an influence on debates, making certain policies visible and obscuring others (Porter and Webb 2007). As a result, PISA is a powerful force for policy convergence around favoured OECD education policies of measurement indicators, basic skills, performance management and accountability. (Rizvi and Lingard, 2006, Porter and Webb 2007, Steiner-Khamsi 2004, Grek 2009).

#### **2.4.7 Neoliberal ideology**

Neoliberalism is the label commonly given to the economic paradigm that has dominated policy making in many countries for the last 40 years. Harvey (2005) sums it up as

A theory of political economic practices that proposes that human well-being can best be advanced by liberating individual entrepreneurial freedoms and skills within an institutional framework characterised by strong private property rights, free markets and free trade (p.2)

It strongly favours the provision of social needs by private interests rather than publicly-run services and believes that competitive markets ensure this happens most efficiently. It favours low taxes, particularly on business, and is vehemently opposed to organised labour, which prevents workers being paid a market rate. The two key neoliberal theorists, Nobel-prize winning economists Fredrich Hayek and Milton Friedman, had been developing their ideas in the post-World War 2 decades. After the economic crisis of the 1970s, neoliberalism replaced Keynesianism as the dominant economic paradigm in developed countries and within the major international organisations including the World Bank and OECD. (Turner 2008, Grant 2009, Crouch 2011, Robertson 2008). Neoliberalism guided the economic reforms of the right of centre Thatcher and Reagan Governments in the early 1980s, and has subsequently set the policy framework in most developed countries, often with only minor variations depending on the party holding power.

Neoliberal theorists and advocates have well developed views on education. In *Capitalism and Freedom* (1962) Friedman wrote of the need for schools to be subject to competition by providing parents with vouchers which could be spent at any school. His schools market would encourage diversity, generate efficiency and pay teachers according to ‘merit’ and market forces. Ravitch (2010) writes that Friedman’s writing underpins the widespread support for school choice in the United States. Kumar and Hill (2008) describe the neoliberal plan for education as marked by:

- Markets and ‘parental choice’
- Privatisation
- Cutting state subsidies.

Writing of the United States context, Hursh (2007) comments, ‘Neoliberal ideals, although rarely explicitly stated, form the basis for most of the education reform proposals since *A Nation at Risk*, published in 1983.’

The alignment of neoliberal ideas with powerful interests is key to their success. The spread of neoliberal ideas and policies around the globe has been aided by pro-free market think-tanks or institutes active in many countries. Usually funded by supportive wealthy individuals and business interests, they intervene in public debate advocating deregulation, privatisation and reducing workers’ bargaining rights. They also help

marshal political support for parties and individuals who subscribe to these views. These organisations are skilled at getting their views aired in mostly supportive major media outlets. Their efforts create a sympathetic environment for neoliberal ideas promoted in transnational policy networks to be introduced locally Steiner-Khamsi (2004) writes that ‘for transfer to succeed, groups of local protagonists benefitting from, believing in, and advocating a specific policy agenda are indispensable’ (p. 211) and neoliberal think tanks and institutes certainly provide these.

In the United States, business-funded institutes like the Gates, Walton and Heritage foundations contribute around \$100 million per year into organisations advocating for vouchers and school choice (Gutstein 2010). The ideologically-driven nature of free market advocacy organisations means they act in accordance with a firmly established world view with well developed ideas on schooling and are not swayed from these by countervailing arguments or evidence. After voucher proposals failed to win wide support in the United States, the establishment of privately run charter schools became the vehicle for expanding school choice in the US and, according to Ravitch (2010), the push for their creation ‘has become more than a reform, it has become a movement’ with support far outstripping the supporting evidence.(p.146).

Allied think tanks, academics, business representatives, private media and members of the public that support neoliberalism can be said to comprise a movement. Indeed it is mostly due to its intersection with neoliberalism that the global education reform movement has the character of a movement. The influential interventions in pursuit of an ideologically-inspired free market agenda in the face of opposition and conflicting evidence that mark broader economic and policy debates are echoed in education.

## **2.5 Summary**

Convergence of education policies in the era of globalisation is driven by a variety of forces. Particular policies of the global education reform movement can be seen to result from more than one of these. For example, the shift toward local autonomy for schools can be seen as a response to increased demands from parents for more say in their children’s education but also as driven by the neoliberal objective of reducing the state’s responsibility for education provision. Generation of education indicators based on student tests are often politically popular because they meet a desire for quantitative

evidence of school effectiveness, yet they also provide the information needed to enable a schools market to function, an official policy preference of the World Bank.

The global trend towards similar education policies in the late 20<sup>th</sup>/early 21<sup>st</sup> century is best understood as the net influence of the different actors and influences outlined above.

### **3. DISCUSSION OF RESEARCH METHODS**

#### **3.1 Quantitative non-experimental research**

PISA conforms with Gall, Gall and Borg's (2010) definition of quantitative research as being based on "epistemological belief in an objective reality, the analysis of reality into measurable variables, the creation of generalisable knowledge through the study of samples that accurately represent a defined population, and reliance on statistical methods to analyse data" (p.124). Student performance and contextual information are quantified to facilitate their analysis using descriptive and inferential statistical methods.

PISA and the current study are both non-experimental research. In experimental research the researcher manipulates an independent variable to measure its effects on two or more groups of subjects - randomly assigned to control for extraneous variables - including a control group which does not receive the treatment (Lauer, 2006). Experimental research can provide much stronger evidence for causality, or the effect of the independent variable on the dependent variable. (Johnson and Christensen, 2000). Non-experimental research is that in which independent variables cannot be manipulated and/or their manifestations have already occurred (Kerlinger and Lee, 2000). Non-experimental research is important to the study of education as it is commonly not possible or ethical to manipulate a variable and phenomena must be observed as they occur in their real world context.

Relationships observed in non-experimental research should be considered with caution or in Gorard's words "causation should not be inferred from correlation" (2001, p.153). The findings of non-experimental research are usually presented in qualified terms and any action recommended by the researcher based on their finding is commonly limited to further investigation (Johnson, 2000). Such caution is required as the problem of alternative explanations for, or the effect of third variables on, observed relationships is omnipresent in non-experimental research (Johnson, 2001, Gorard 2001). The OECD acknowledges the limitations that can be placed on the interpretation of data PISA collects from schools and their relationship to student performance, warning

The learning environment in which 15-year olds find themselves and which PISA examines may only be partially indicative of the learning environment that shaped

their educational experiences....To the extent that the current learning environment of 15-year-olds differs from that of their earlier school years, the contextual data collected by PISA is an imperfect proxy for the cumulative learning environments of students, and their effect on learning outcomes is therefore likely to be underestimated. (OECD, 2007b, p.215)

Introducing their international comparison of equity in education systems which utilised data collected through PISA, Gorard and Smith (2004) write that while it can be very useful

Comparison between different national education systems is difficult...largely because school systems differ in so many respects that it is difficult to mount a convincing argument that any one difference is related to another. (p15)

Such reservations are often absent in public policy debates in which PISA findings are referenced.

The risk of erroneous interpretations is heightened where research is not guided by testing of hypotheses and where relationships are observed and *post hoc* the researcher derives an explanation (Kerlinger and Lee, 2000, Johnson and Christensen 2000).

### **3.2 Correlational research**

Two common types of non-experimental research are causal-comparative and correlational studies. They are both used to study relationships between independent and dependent variables in real world settings. Where they differ is in the measurement of variables. Causal-comparative studies are used where one or more variable is classified categorically and correlational where you have two continuously scaled quantitative variables (Johnson 2001). In its analysis of the PISA data, the OECD uses comparative and correlational methods. An example of the former is PISA's 2006 analysis of the effects of school competition in which schools were designated as either competing with at least one other school to enrol students or not. A comparison of mean science scores found schools subject to competition to have on average a score 17.9 % higher, which was statistically significant. In 2009 PISA used cross-country correlations to look at relationships between percentages of students in schools implementing a variety of accountability, governance and admittance policies and mean reading scores. The current

study is correlational as PISA measures student achievement and contextual information on implementation of GERM policies at the national level on continuous scales.

Johnson (2001) counters the ‘misleading belief’ (p.4) that causal comparative research provides superior evidence of causality than correlational research when the only difference is the scaling of the independent and/or dependent variable and in both evidence for causality is practically non-existent.

Correlational studies entail primarily the computing of correlation coefficients to explain the strength of the relationship between two variables. This is most commonly – and in the case of this study – done using the Pearson product-moment correlation coefficient ( $r$ ) (Gall, Gall and Borg 2010). This has a value between + and -1 and the closer to + and -1  $r$  is, the stronger the correlation. A positive correlation means that as the amount of  $x$  increases, so does  $y$ . For a negative correlation, as  $x$  increases,  $y$  decreases. A correlation coefficient of 0 means there is no relationship between variables.

Once arrived at, it is necessary to determine if a correlation is statistically significant, that its value is unlikely enough to enable rejection of the null hypothesis. A result is statistically significant when the evidence suggests it was probably not due to chance. To determine this it is necessary to set a significance level ( $\alpha$ ) (Gorard, 2001). In educational research, this is commonly set at .05, or, results with a less than 5 per cent probability they were due to chance are considered statistically significant. When using correlations in its analysis of its survey data, PISA generally identifies results that are significant at the .05 level, although in a small number of cases uses a significance level of .1.

### **3.3 Secondary analysis**

Cook (1974) defines secondary analysis as “attempts to use existing data from basic research or evaluations to assess the degree of empirical support for major assumptions underlying present practices or alternatives for future policy” (p.159). Hakim (1982) describes secondary analysis as “any further analysis of an existing dataset which presents interpretations, conclusions, or knowledge additional to, or different from, those presented in the first report on the inquiry as a whole and its main results” (p.1). The current study, reanalysing data collected as part of PISA, fits these descriptions of secondary analysis.

Secondary analysis has a well established place within educational research. The landmark 1966 Coleman Report *Equality in Education* found that differences among United States schools accounted for only a small fraction of differences in student achievement and that students' background was much more closely related to achievement. This led to many policy changes including school bussing (Kerlinger and Lee 2000). Subsequently the report's extensive data on verbal achievement and over 100 independent variables was subject to extensive reanalysis by different parties, some of whom had had their assumptions challenged. This reanalysis took different forms including questioning the Coleman methodology and using sub-samples of the data to approach similar questions from different approaches. While some concerns were raised with the Report's methodology, its findings were corroborated (Smith, 2006). This process was an example of secondary analysis applying rigour to, and ultimately strengthening, original research.

Secondary analysis can constitute original research (Gorard, 2001) and it has an important role in an era of large data gathering surveys such as PISA and the Trends in International Mathematics and Science Study (TIMSS). Secondary analysis is valuable for the new perspectives and freedom it can bring to analysing and rephrasing questions and asking new ones. It can also bring role independence free of any explicit or implicit influence the funder of the original research may have on what is done with the data, which questions are asked etc. (Cook, 1974). The availability of PISA data since the first survey in 2000 has led to greater secondary analysis among educational researchers. Examples include Gorard and Smith's (2004) use of PISA contextual data to investigate school segregation in Europe, Smith's (2001) analysis of PISA results of low-income background students to refute the view that Britain has a long tail of underachievement and Gorard, See, Smith and White's (2006) use of information from the principals survey to investigate patterns of teacher supply.

Secondary analysis is not without its pitfalls. Cook (1974) is critical of Hanushek and Kain's use of Coleman Report data on negro twelfth graders in northern US states. While they reached findings that differed from those in the original report, Cook says this is due to sampling bias, or they selected a sub-sample of the data that gave them a dramatically different result from the report's general finding. Cook states that such sampling is suspect when no compelling reason for it is advanced.

There is also the criticism that original data used in secondary analysis may not be value neutral. How survey questions were phrased and even what questions weren't asked are relevant matters and the secondary analyst needs to interpret data with a good understanding of how and why the original data collection was carried out (Dale, Arber and Procter, 1988).

When engaging in secondary analysis then it is necessary to understand

- the original survey's aims and sponsors,
- procedures for data collection
- sampling design and response rates
- how data was collected for the variables
- how variables were defined, calculated and quantified (Smith, 2001).

### **3.4 Appropriateness of current study, methods used and research questions**

Aware of the global influence of PISA on education policy, I decided to subject its findings to secondary analysis after reading the executive summary that accompanied the release of the 2006 results. This contained key findings across a broad range of school organisation matters, including three that supported GERM policies. The 2006 Analysis volume explained that these findings had been arrived at through varying statistical methods using data at the school and national levels. Looking through PISA's 2006 Data volume and data made available on the PISA website I realised it would be possible to conduct correlational analysis at the national level of the relationship between the degree of implementation of GERM policies and student achievement (OECD 2007a, 2007b, 2007c). This would involve the use of a consistent statistical method to assess the relationship of a number of variables with student learning outcomes. This would involve looking at some of the relationships PISA had already investigated using different statistical methods and data collected at the school, not national, level. Given the timing of my study I was able to repeat the analysis using PISA 2009 data, noting that PISA subsequently adopted the cross-country correlation method I had first used with 2006 data.

The decision to use PISA data is taken in full awareness of the warnings the OECD attaches to it: that the learning environment of a 15 year old at the time they undertake

PISA is not necessarily the same as that which shaped earlier learning and as such the contextual data collected is an imperfect proxy for the cumulative learning environments of students, and that contextual data on school system characteristics is gathered from answers to the principal questionnaire which received an average of 300 responses per country in 2006 and 264 per country in 2009. Despite its limitations and critics, PISA is the most comprehensive source of internationally comparable data on education systems and learning outcomes and employs high levels of technical expertise and detailed quality control measures.

The issue of the strengths and weaknesses of PISA data is an important one, and one that I discuss, but in a way separate from the need to conduct secondary analysis of it. Backed up by the institutional power of the OECD, PISA already occupies an important place in international education policy formulation and its findings are influential. Secondary analysis like the current study bring important scrutiny to PISA's work and findings, and identify relationships between school system characteristics and learning outcomes, or the absence of them, that PISA has not identified. It can also help better understand the role of the OECD and PISA and their relationship to the global education reform movement.

## **4. PISA OPERATIONS AND DATA COLLECTION**

### **4.1 Background**

The Programme for International Student Assessment (PISA) has been administered every three years since 2000 by the Organisation for Economic Cooperation and Development (OECD). Randomly sampled students complete a two-hour test designed to assess their competency in science, mathematics and reading literacy. PISA claims not to test students' retention of curriculum based knowledge but instead assesses "in terms of the acquisition of broad concepts and skills that allow knowledge to be applied" (OECD 2006 p.11).

57 nations/education systems participated in PISA 2006 comprising all 30 OECD member countries and 27 partner countries. The OECD claimed these countries represent almost 90 per cent of the world's economy (OECD, 2007a). Around 400,000 students participated in the survey representing approximately 20 million students in the eligible age bracket of 15 years 3 months to 16 years 2 months at the time of assessment. Participating countries in 2006 and 2009 had to conduct the surveys within a 42 day period between March 1 and August 31. For PISA 2009 participation expanded to around 470 000 students in 65 countries/education systems representing about 26 million school-goers in the target age group. (OECD 2010a) This age bracket is selected to provide a sample of students who have completed at least six years of formal schooling regardless of their system's policy on starting age (OECD, 2006). PISA believes it "monitor(s) the outcomes of education systems in terms of student performance" and "assesses the extent to which students near the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in society" (OECD, 2007b pp. 3, 16).

The term 'countries/education systems' is used above because although most participating countries sample students from their entire eligible population, China only does so in three discrete geographical and administrative areas: Hong Kong, Macao and Shanghai (2009 only). Throughout, when participating countries are

referred to collectively this includes these sub-national Chinese systems as distinct entities (OECD 2007b, OECD 2011).

PISA also collects extensive detailed contextual information from participating students and schools. Some of the information collected in the survey of principals of participating schools is done so to facilitate analysis of the relationship between school organisation policies and student learning outcomes. Some of the school accountability, autonomy and competition policies on which data is gathered are central to the global education reform movement

PISA's test results and collection of contextual data provide the opportunity to assess performance between countries and to investigate relationships between performance and school and social contexts within and between countries. The OECD carries out substantial analysis using the PISA data. For PISA 2006 this analysis was published in one volume and for 2009 it was spread across six volumes. The data is also made public through the PISA website for use in secondary analysis, such as the current study.

#### **4.2 Administration**

PISA activity is directed by a governing board (PGB) comprising representatives of all OECD member nations. The task of designing and implementing PISA surveys is passed to an international consortium of educational research bodies. The PISA 2006 consortium was led by the Australian Centre for Educational Research (ACER) and also comprised the National Institute for Educational Measurement (CITO) in the Netherlands, WESTAT and the Educational Testing Service (ETS) in the United States, and the National Institute for Educational Policy Research (NIER) in Japan (OECD 2006). Responsibility for PISA 2009 was shared between two consortia. One led by CITO and also including staff from the University of Twente (Netherlands) Behavioural Science Faculty, University of Jyväskylä (Finland) Institute for Educational Research and the French Ministry of Education, was responsible for the contextual questionnaires. The other consortium led by ACER was responsible for all other aspects of PISA 2009 data collection (OECD 2011)

Development of the scientific literacy framework and other requirements for PISA 2006 began in December 2002 when the OECD invited international experts in science education to a forum. From there development of the framework was carried out by a

working group, later superseded by the Science Framework Expansion Committee, and finally by the 2006 PISA consortium once the contracts were let. Similar processes were followed for reading and mathematical literacy ahead of the 2000 and 2003 surveys (OECD, 2009). The 2006 PISA report was subtitled ‘Science Competencies for Tomorrow’s World’ and the majority of questions tested students’ scientific literacy (54%). Reading and mathematical literacy were also examined (15% and 31% respectively). Of the 108 science test items used in 2006, 22 were developed for and used in PISA 2003. New items were tested in field trials in 2005. PISA 2006 used reading and mathematics items developed for the 2000 and 2003 surveys (OECD, 2009).

For 2009 PISA returned its focus to reading as it had been for the initial survey in 2000. Development of the reading literacy framework to be used began in August 2006. This process, involving test developers from the ACER-led consortium, retained much of the substance of the PISA 2000 but added new elements including digital reading assessment using computers. The 131 reading items comprised 37 that had been used in earlier years and 94 that were developed for PISA 2009 following field trials in all participating countries during 2008. Assessment of mathematics (34 items) and Science (53 items) used items from 2006 (OECD 2011).

In each country PISA appoints a national project manager (NPM). Working from a national centre (NC), the NPM implements PISA procedures and coordinates school level activities with school co-ordinators (SCs), usually a staff member. PISA is overseen within schools by test administrators (TAs), who cannot be science, reading or mathematics teachers of students being assessed and preferably not staff members of any participating school. For PISA 2009 in some countries the roles of the SC and TA were undertaken by one person referred to as the school associate (OECD 2009, 2011).

The test itself lasts two hours and students receive one of thirteen different questionnaire booklets. Tests are required to be administered to students in a uniform manner to the point of the introduction and explanation given by the TA being read from a prescribed text. Test booklets, questionnaires and other materials are sent to SCs in sealed packages and must remain secure until the test session. Upon completion of the test TAs are responsible for returning all test materials to the NC where numbers of completed and unused booklets were checked against participation status information recorded by the TA (OECD 2009, 2011).

### 4.3 Translation and cultural appropriateness

PISA is taken in each participating country in the local language of instruction. This includes some regional languages/dialects spoken among sub-sections of the population. For PISA 2006, 87 national versions of the test materials in 44 languages were used and in 2009, 101 versions in 45 languages. This required translation from the source languages, English and French, in accordance with strict procedures. (OECD 2009, 2011)

Submissions of potential test items are invited from participating countries. These are then circulated for feedback via NCs to assess whether they are relevant to a country's curriculum, would be interesting to students, raised any cultural concerns or be difficult to translate. Following field trials, test items are decided upon by the responsible consortium and source versions of the questionnaire are developed in French and English. Double translation into local languages from both source languages is then performed under the auspices of NPMs. Double translation involves two independent translations from the source and reconciliation by a third party and is preferred by PISA over back translation (translating a source version into the target language, then translating it back and comparing discrepancies) as it requires three different parties to compare the source and target versions, and discrepancies are recorded in the target language. PISA uses two source languages to reduce the impact of syntactical features and cultural characteristics contained within one language. (OECD 2009, 2011)

All translators recruited by NPMs receive *PISA Translation and Adaptation Guidelines* containing detailed instructions on how the process is to be carried out as well as tips on common traps, how to adapt material to the national context and more. All versions underwent international verification except for national versions created for schools teaching students from ethnic minorities that make up less than 5% (2006) or 10% (2009) of the target student population i.e. Irish-Gaelic. As the small numbers of these students will have negligible impact on country results, national level verification is deemed sufficient. Verification involves NPMs submitting their version(s) of the questionnaire, manual and coding guidelines translated from the source versions, to members of the consortia which employ

teams of translators with a command of the target language and professional experience translating between it and English and/or French. Formatting of national versions of the test booklets has to follow as far as possible the layout in the source versions (OECD 2009, 2011).

#### **4.4 Quality assurance**

All persons involved in the implementation of PISA are required to perform their function in line with the official operational manuals:

- PISA National Project Manager's Manual
- PISA Test Administrator's Manual
- PISA School Coordinator's Manual,
- PISA School Sampling Preparation Manual,
- PISA Data Management Manual.

National level planning documents are developed by each country from the operational manuals and their content has to be agreed between the NPM and the consortium before being posted on the PISA website (known as the MyPISA website from 2009). Before printing, assessment materials for each country are subject to a final optical check by the consortium. This aspect of PISA's QA process is continually being refined and for 2009 new spreadsheets and report forms were introduced at this stage to identify issues for follow up and identify whether corrections identified earlier had been implemented. Despite this, students' answers to a number of items on particular national versions of the test had to be omitted for the computation of national scores for reasons including mistranslations, confusing translations and poor printing. (OECD 2009, 2011)

In 2006 National Centre Quality Monitors (NCQMs) appointed by the consortium visited all 57 participating national centres in the month preceding the country's testing period. In 2009 this visiting was limited with a focus on newly participating countries. As well as advising national centre staff, NCQMs conduct structured interviews with NPMs covering all areas of administration of PISA in their country. The results of these are recorded. PISA Quality Monitors (PQMs) are

employed by the consortia and based in participating countries. They visit a sample of schools (15 per country in 2006, 7 or 8 in 2009) mostly unannounced, to check and record that all aspects of test administration at this level are carried out correctly (OECD 2009, 2011).

#### **4.5 Sampling**

The national target population from which PISA draws its samples is all students in the target age bracket (15 years 3 months to 16 years 2 months) who are enrolled in an educational institution. A sample design and size is then determined for each country to, in PISA's words, "maximise sampling efficiency for student-level estimates" (OECD 2007b p. 24). Sample sizes are not a simple function of the size of a country's target population, but derived from calculations of the design effect on standard error introduced by measurement error and PISA's complex sampling design.

The PISA sample population is sampled in clusters (schools). As students attending the same school have had similar education and are likely to come from similar backgrounds, standard error is introduced relative to a randomly chosen sample. This is known as the design effect. Stratified sampling of schools and selection of schools with probability based on their size are used to help limit this standard error but cannot eliminate it. For this reason PISA calculates design effects introduced by the between school variance and measurement error for participating countries, and from these, effective sample sizes. A country's sample size should be equal to or greater the largest of the effective sample sizes calculated for the different design effects PISA assesses (OECD 2009).

This explains why two countries with similar target populations required markedly different sample sizes. The greater the variance between schools, the greater the design effect and larger the effective sample size has to be. In 2006 Sweden with a target student population of 127,036 sampled 4,443 students while the Czech Republic with a target of population of 122,764 sampled 5,932 students. For reading, the design effect, or inflation of the total variance due to measurement error and complex sampling design, was 5.44 for Sweden and 8.38 for the Czech Republic (OECD 2009).

The target school population in each country is stratified prior to sampling to improve sampling efficiency, make sure all parts of a population are included in the sample and ensure adequate representation of specific groups within a population. Stratification variables include school size, public/private, school type (i.e. academic, vocational) and urban/rural. A minimum of 150 schools has to be selected in each country or all schools if there are less than 150 eligible. School sampling is conducted by the PISA consortium from a sampling frame of all schools containing eligible students supplied by NPMs and stratified according to agreed criteria. Once a school is sampled the SC sends a list of all eligible 15-year-old students to the NPM. From this list, students are randomly selected to fulfil the target cluster size (TCS), typically 35, are randomly selected, or all of those in the target age bracket if the school had less eligible students than the TCS (OECD 2009, 2011).

#### **4.6 Maintaining precision in the sample**

Exclusions from the target population are only permitted in limited circumstances with the aim of limiting them to 5 per cent. This seeks to “ensure that under reasonable assumptions any distortions in national mean scores would remain within plus or minus 5 score points” (OECD 2007b, p.23). Exclusions at the school level are permitted where remoteness, size or operational factors would create unreasonable difficulty. Schools enrolling only students with special educational needs could be excluded, up to a maximum of 2%. If a country’s proportion of such schools is higher than this the schools have to remain in the sampling frame and if selected, their students complete the PISA UH booklet containing a subset of questions more suitable. Within schools students can be excluded on the basis of intellectual disability, physical incapacity, psychological unsuitability or limited proficiency in the language of the test. For PISA 2006, the overall exclusion rate was below 4% in 51 of the 57 countries and below 6% in all countries except Denmark (6.07%) and Canada (6.35%) (OECD 2007b). For PISA 2009, the overall exclusion rate was below 4% in 53 of the 65 countries and below 6% in all countries except Canada (6%), Luxembourg (8.15%) and Denmark (8.17%) (OECD 2011 Table 11.1).

Within a country 85% of selected schools are required to take part, although replacement schools are allowed if a selected school does not respond. Schools in which less than 25% of selected students participate are regarded as non-respondents and the data from students not used. Schools with a student participation rate between 25 and 50% are not considered participating schools but the data from students in these schools is retained and used. This formulation was arrived at as a balance between, a) non-response bias arising from the extent to which students participating at a school differ in achievement from those not attending, b) non-response bias resulting from excluding an entire school to the extent that an individual school was different from others in a country, and, c) increased sampling variance resulting from a smaller sample size. A response rate of 80% of all selected students within a country is required. (OECD 2009, 2011)

For PISA 2006 weighted school response rates after replacement were 100 per cent in some cases and over 95 per cent in a large majority of cases. A smaller number of lower school response rates were observed notably the United States (79.09 per cent) Canada (86.23 per cent) and United Kingdom (88.15 per cent). Weighted student response rates after replacement ranged from 99.55 per cent (Azerbaijan) to 81.43 (Canada) (OECD 2009). For PISA 2009 weighted school response rates after replacement were above 85% in all cases except the United States (77.5%) and Panama (83.8%). After undertaking ‘additional analysis’ PISA determined that ‘no notable bias would result and the data from both was included. Weighted student response in 2009 was above 80 % in every country except Canada (79.5%). Again PISA conducted ‘additional analysis’ and found no notable bias would result and the data from Canadian students was retained in the final database (OECD 2011).

#### **4.7 The PISA test, coding and national mean scores**

The PISA test lasts two hours. Each student receives one of 13 different questionnaire booklets with no more than three students in any school group of 35 receiving the same booklet. Item formats vary between multiple choice and those requiring a constructed response ranging from a simple word or number to more detailed explanation. Short closed-constructed response questions, require an answer based on a set of provided possible responses, and are assessed as either incorrect or correct. For PISA 2006 the 13

different questionnaire booklets were designed such that science questions comprised 54% of all assessment, mathematics 31%, and reading 15%. This was achieved by assigning questions to one of thirteen item clusters (seven science, four mathematics and two reading). Each booklet comprised four item clusters. In 2009 the booklets were comprised of four clusters assigned from seven reading clusters, three maths clusters and three science clusters(OECD 2009, 2011)

#### **4.7.1 Science**

PISA developed a comprehensive scientific literacy framework for the 2006 survey and summarises it thus: “scientific literacy requires an understanding of scientific concepts, as well as the ability to apply a scientific perspective and to think scientifically about evidence” (OECD 2007b, p.21). This definition of scientific literacy develops that used in the 2000 and 2003 surveys by including attitudinal responses and increased emphasis on the nature and methodology of science. While attitudinal science items requiring students to express their interest in or support for science were included, students were told there were no correct answers and they would not count towards their test scores (OECD 2009). Science questions were framed so as to be relevant and be posited within contexts students are familiar with namely: health, natural resources, environmental quality, hazards and frontiers of science. The three competencies required of students were the ability to identify scientific issues, explain phenomena scientifically and use scientific evidence. Questions were designed to test both knowledge of science and about science and were drawn from the major science fields of physics, chemistry, biology, Earth and space science, and technology (OECD 2007b). In PISA a unit begins with a text and/or visual stimulus, which is followed by a number of questions. Box 1 contains an actual science unit from PISA 2006.

## Box 1.

Science unit from PISA 2006

The Greenhouse Effect: Fact or Fiction?

Living things need energy to survive. The energy that sustains life on the Earth comes from the Sun, which radiates energy into space because it is so hot. A tiny proportion of this energy reaches the Earth.

The Earth's atmosphere acts like a protective blanket over the surface of our planet, preventing the variations in temperature that would exist in an airless world.

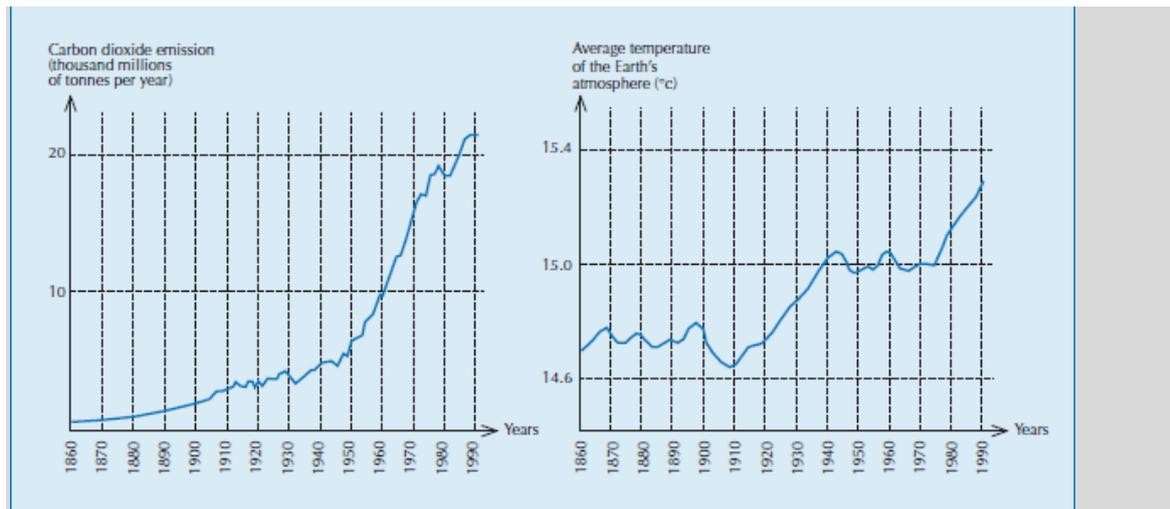
Most of the radiated energy coming from the Sun passes through the Earth's atmosphere. The Earth absorbs some of this energy, and some is reflected back from the Earth's surface. Part of this reflected energy is absorbed by the atmosphere.

As a result of this the average temperature above the Earth's surface is higher than it would be if there were no atmosphere. The Earth's atmosphere has the same effect as a greenhouse, hence the term greenhouse effect.

The greenhouse effect is said to have become more pronounced during the twentieth century.

It is a fact that the average temperature of the Earth's atmosphere has increased. In newspapers and periodicals the increased carbon dioxide emission is often stated as the main source of the temperature rise in the twentieth century.

A student named André becomes interested in the possible relationship between the average temperature of the Earth's atmosphere and the carbon dioxide emissions on the Earth. In a library he comes across the following two graphs.



André concludes from these two graphs that it is certain that the increase in the average temperature of the Earth's atmosphere is due to the increase in the carbon dioxide emission.

#### Question 3

What is about the graphs that supports André's conclusion?

#### Question 4

Another student, Jeanne, disagrees with André's conclusion. She compares the two graphs and says that some parts of the graphs do not support his conclusion.

Give an example of a part of the graphs that does not support André's conclusion. Explain your answer.

#### Question 5

André persists in his conclusion that the average temperature rise of the Earth's atmosphere is caused by the increase in the carbon dioxide emission. But Jeanne thinks that his conclusion is premature. She says: "Before accepting this conclusion you must be sure that other factors that could influence the greenhouse effect are constant".

Name one of the factors that Jeanne means.

All the questions in this unit required what PISA calls an open-constructed response. For Question 3, a full credit was awarded for referring to the increases in both CO<sub>2</sub> emissions and temperatures or referring to a positive relationship between the two. For Question 4, a full credit was awarded for referring to a particular part of the graphs where both are not descending and climbing and giving the correct description i.e. between 1940 and 1975 the temperature stays about the same but CO<sub>2</sub> emissions continue to rise sharply. A partial credit was given in a number of circumstances such as mentioning a period that illustrates the point but without any explanation. For Question 5, a full credit was given for referring to radiation from the sun being a possible factor, referring to natural phenomena such as water vapour or clouds, of the effect of man-made pollutants (OECD 2007b).

For the 2009 survey, PISA used the science framework developed for 2006 and selected test items from those used in 2006 (OECD 2011).

#### **4.7.2 Proficiency levels and question difficulty**

PISA's assessment of a student's level of proficiency is based on a scale of performance, which for science assessment is known as the science scale. All science questions have a position on this scale depending on their degree of difficulty. This allows each student to be given a score which is based on the highest level at which they could be expected to answer correctly a majority of the time. The scale has six levels and is constructed to have a mean score of 500 and a standard deviation of 100 for OECD countries (OECD 2007b). The six levels of the science scale correspond to levels of proficiency in the competencies outlined in the science framework. Table 1 details the six levels of the science scale developed for PISA 2006 and what students achieving that level can typically do.

Table 1.

Summary descriptions of the six proficiency levels on the PISA science scale

<b>Level</b>	<b>Lower score limit</b>	<b>Percentage of students able to perform tasks at each level or above (OECD average)</b>	<b>What students can typically do</b>
<b>6</b>	707.9	1.3% of students across the OECD can perform tasks at Level 6 on the science scale	<p>At Level 6, students can consistently identify, explain and apply scientific knowledge and knowledge about science in a variety of complex life situations. They can link different information sources and explanations and use evidence from those sources to justify decisions.</p> <p>They clearly and consistently demonstrate advanced scientific thinking and reasoning, and they demonstrate willingness to use their scientific understanding in support of solutions to unfamiliar scientific and technological situations. Students at this level can use scientific knowledge and develop arguments in support of recommendations and decisions that centre on personal, social or global situations.</p>
<b>5</b>	633.3	9.0% of students across the OECD can perform tasks at least at Level 5 on the science scale	<p>At Level 5, students can identify the scientific components of many complex life situations, apply both scientific concepts and knowledge about science to these situations, and can compare, select and evaluate appropriate scientific evidence for responding to life situations.</p> <p>Students at this level can use well-developed inquiry abilities, link knowledge appropriately and bring critical insights to situations. They can construct explanations based on evidence and arguments based on their critical analysis.</p>

<b>4</b>	558.7	29.3% of students across the OECD can perform tasks at least at Level 4 on the science scale	At Level 4, students can work effectively with situations and issues that may involve explicit phenomena requiring them to make inferences about the role of science or technology. They can select and integrate explanations from different disciplines of science or technology and link those explanations directly to aspects of life situations. Students at this level can reflect on their actions and they can communicate decisions using scientific knowledge and evidence.
<b>3</b>	484.1	56.7% of students across the OECD can perform tasks at least at Level 3 on the science scale	At Level 3, students can identify clearly described scientific issues in a range of contexts. They can select facts and knowledge to explain phenomena and apply simple models or inquiry strategies. Students at this level can interpret and use scientific concepts from different disciplines and can apply them directly. They can develop short statements using facts and make decisions based on scientific knowledge.
<b>2</b>	409.5	80.8% of students across the OECD can perform tasks at least at Level 2 on the science scale	At Level 2, students have adequate scientific knowledge to provide possible explanations in familiar contexts or draw conclusions based on simple investigations. They are capable of direct reasoning and making literal interpretations of the results of scientific inquiry or technological problem solving.
<b>1</b>	334.9	94.8% of students across the OECD can perform tasks at least at Level 1 on the science scale	At Level 1, students have such a limited scientific knowledge that it can only be applied to a few, familiar situations. They can present scientific explanations that are obvious and that follow explicitly from given evidence.

For Question 3 in the above Greenhouse Unit, a full credit answer was awarded 529 points, as such it was a Level 3 question. For question 4 a full credit was awarded 659 points and a partial credit 568 points. Full credit answers to the more difficult (Level 6) Question 5 received 709 points. The relative difficulty of questions is estimated by considering the proportion of students getting each question correct (OECD 2007b). PISA's proficiency scale gives a picture of the spread of proficiency levels among the student population of a country. It is possible to estimate the location of individual students on the scale, giving them a score according to the hardest task they could be predicted to perform with a given probability. Students are placed on the scale at the point where they have a 62 per cent chance of correctly answering questions located at the same point. (OECD 2009)

### **4.7.3 Reading**

Based on the reading literacy framework developed for PISA 2000, the reading literacy competencies tested for are the ability to:

- Retrieve information
- Interpret texts, and
- Reflect on and evaluate texts. (OECD 2007b).

With a return to reading literacy as the major domain for 2009, PISA revisited the framework, expanding it to incorporate digital reading and metacognition in reading, or what PISA refers to as reading strategies. Of the 131 reading items used in PISA 2009 94 were newly developed with the remainder having been used in earlier surveys. Also, some countries that had achieved a national mean reading score of below 450 in PISA 2006 were offered the option of using booklets that contained easier reading items to obtain better descriptive information about what students can do and make the test more satisfying for students. This was done by replacing two of the standard reading clusters with ones containing the easier items (OECD 2011).

PISA's combined reading proficiency scale has five proficiency levels with a mean score of 500 points and standard deviation of 100 (OECD 2007b). A separate proficiency scale

was used for the digital reading assessment introduced for PISA 2009. Box 2 contains an example of a multiple choice answer reading literacy question from PISA 2006. Answering correctly - A: About two metres- earned a full credit. The assessed level of difficulty of the question was determined to be 478 points, or Level 2 on the reading literacy scale.

Box 2.

Reading Unit from PISA 2006

### Lake Chad

Diagram A shows changing levels of Lake Chad, in Saharan North Africa. Lake Chad disappeared completely in about 20000 BC, during the last Ice Age. In about 11000 BC it reappeared. Today, its level is about the same as it was in AD 1000.

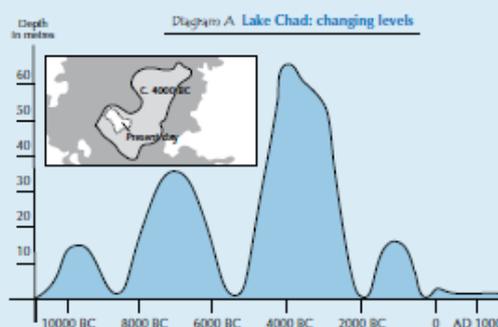
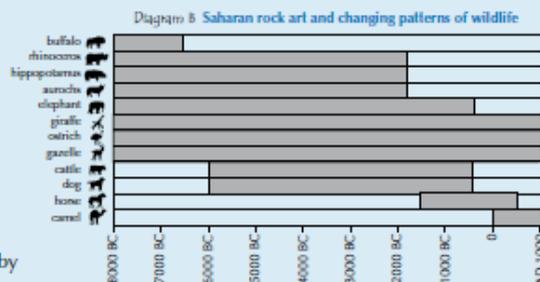


Diagram B shows Saharan rock art (ancient drawings or paintings found on the walls of caves) and changing patterns of wildlife.



Source: Copyright Bartholomew Ltd 1988. Extracted from *The Times Atlas of Archaeology* and reproduced by permission of Harper Collins Publishers.

What is the depth of Lake Chad today?

- A. About two metres.
- B. About fifteen metres.
- C. About fifty metres.
- D. It has disappeared completely.
- E. The information is not provided.

(OECD 2007b, p.290)

#### 4.7.4 Mathematics

PISA defines mathematical literacy as “an individual’s capacity to identify and understand the role that mathematics plays in the world, to make well founded judgements and to use and engage with mathematics in ways that meet the needs of that individual’s life as a constructive, concerned and reflective citizen” (OECD 2007b, p.304). As with science and reading, PISA tests students’ literacy in mathematics in an applied context and is not simply concerned with whether they can perform basic maths functions. The scale used to assess question difficulty/student proficiency in mathematics has six levels and a mean of 500 points with a standard deviation of 100. This scale and the framework for assessing mathematical literacy were developed ahead of PISA 2003 which had mathematical literacy as its focus (OECD 2009). Box 3 is an example of a PISA mathematics question requiring a short-constructed response. A full credit for a correct answer – 12,600 ZAR – was worth 406 points putting it at Level 1 on the mathematics scale. PISA 2009 used the mathematics framework devised for the 2003 survey and the mathematics items were selected from those used in 2006 (OECD 2011).

Box 3.

Mathematics Unit from PISA 2006

##### EXCHANGE RATE

Mei-Ling from Singapore was preparing to go to South Africa for 3 months as an exchange student. She needed to change some Singapore dollars (SGD) into South African rand (ZAR).

Mei-Ling found out that the exchange rate between Singapore dollars and South African rand was:

$$1 \text{ SGD} = 4.2 \text{ ZAR}$$

Mei-Ling changed 3000 Singapore dollars into South African rand at this exchange rate.

How much money in South African rand did Mei-Ling get?

(OECD 2007b, p.311)

#### 4.7.5 Coding of Students' Answers

Uniformity in coding student responses to open-ended items is crucial to PISA's validity and considerable lengths are gone to to achieve this. NPMs have to translate national coding guides and submit them to the consortium for verification. They then have to recruit and train coders. These require a good understanding of either mid-secondary level mathematics or science or the language of the test and need to be familiar with ways in which secondary level students express themselves. A minimum of four coders is required in each of the three domains although PISA now recommends having 16 coders for the major domain. Full-time supervisors and leaders monitored coding teams for reliability and consistency and PISA believes their role is essential as they monitor coders' consistency in applying the coding criteria. NPMs have access to a query service on the PISA internal website for questions not covered in the coding manual and which cannot be resolved at the national level. All coders must undertake training that involves reviewing the coding for the actual clusters used in the test and assigning of codes to sample answers (OECD 2009, 2011).

Completed test booklets are randomly allocated to coders. Each test booklet contains four clusters. A coder is allocated a cluster and then codes it in the four different booklets before moving on to another cluster. Further, each individual item within a cluster has to be coded across the different booklets before moving on to the next item. The design for allocating clusters to coders means each student's booklet is coded by four different coders (one for each cluster), a way of minimizing the effects of any systematic leniency or harshness (OECD 2009, 2011).

For PISA 2009 each country also had to set aside 100 of each booklet for multiple coding of items in the first cluster by four coders to check for consistency. In 2006 this was carried out on six of the booklets by four coders. Cross-national coding consistency was checked in 2006 by requiring NPMs to submit a sub-sample of responses for an International Coding Review by an independent panel of experts and in 2009 through an inter-country coder reliability study (OECD 2009, 2011).

#### **4.7.6 Weighting**

For the purposes of calculating distributions of students' proficiency levels and national mean student scores it is necessary to weight each student's score so that they represent the correct number of students in the total population. While students are chosen randomly, they do not have the same probability of being chosen. Reasons for this include over representation of schools of particular interest (i.e. ethnic or indigenous groups) in the sample design, schools turning out to be larger or smaller than expected (affecting the probability of being sampled) and non-response. The design could also under sample the number of small or remote schools for operational or cost reasons.

The weight for a student is a product of school base weight (the reciprocal of the probability of inclusion of the school in the sample) and within school base weight (the reciprocal of the probability of selection of a student from within a selected school) and five adjustment factors designed to:

1. Compensate for non-participation by other schools similar in nature to the sampled school (not already compensated for by inclusion of replacement schools)
2. Compensate for some countries/schools only including 15-year-old students enrolled in the modal grade for 15-year-olds.
3. Compensate for non-participation by students within the same school and sampling strata
4. Trim, or reduce, unexpectedly large school base weights
5. Trim, or reduce, weights of students whose weights are exceptionally large (OECD, 2009).

With the correct weightings applied, scores of all participating students within a country can be used to calculate a national mean score for each country.

#### **4.8. Contextual Information**

Following a short break at the end of the two-hour test, students' complete a 30 minute context questionnaire. This asks them to supply the following information about their background:

- Parental occupation and student's expected occupation at age 30.
- Highest educational level attained by their mother and father.
- Immigrant background, defined by PISA as native, second generation or first generation.
- Whether the language of assessment is used at home.
- Availability at home of the following indications of family wealth:
  - A room of their own
  - A link to the internet
  - A DVD/VCR player
  - A dishwasher
- How many of the following they have at their homes:
  - Cellular phones
  - Televisions
  - Computers and cars.
- Students are also asked about the presence of three other country specific items thought to indicate wealth.
- Home Educational Resources: students are asked whether they have a desk to study at, a quiet place to study, a computer they can use for school work, educational software, their own calculator, books to help with their school work and a dictionary (OECD 2007b).

Indices are constructed for these variables with parameters estimated from sub-samples of students from each OECD country and scaled based on computed estimated values on the relevant index for all schools and students. These are then standardised so that the mean of the index value for the OECD student population is zero and the standard deviation one. (OECD, 2007b)

PISA has also created a combined index of economic, social and cultural status (ESCS) to more broadly define student background. This comprises the index of parents' highest educational level (father or mother), index of highest international socioeconomic status of occupational status (father or mother) and the index of home possessions based on the survey of household possessions and educational resources. PISA's rationale for the use of these variables is that socio-economic status is usually seen as being determined by occupational status, education and wealth. As information on parental income is not available, presence of household items is used as a proxy for wealth (OECD, 2007b). The

ESCS is used in PISA's analysis of the impact of socio-economic background on student performance. In 2006 it was found that across OECD countries, an increase in one standard deviation on the ESCS index is associated with a 40 points higher science score (OECD, 2007b). In 2006 PISA also carried out a hypothetical adjustment of mean national science scores assuming the mean ESCS was equal across all OECD countries. This adjusted scores downwards for countries with an above average ESCS and upwards for countries with a below average ESCS (OECD, 2007b). The effect of this adjustment can be seen in Table 2 alongside the national mean science, mathematics and reading student scores for participating countries.

Table 2.

Mean national scores for science, mathematics and reading and adjusted mean science scores if mean economic, socio-economic and cultural status were equal in all OECD countries, all participating countries, PISA 2006

	Mean Science Score	Mean Science Score if the mean ESCS would be equal in all OECD countries	Mean Mathematics Score	Mean Reading Score
Australia <sup>OECD</sup> Nations	527	519	520	513
Austria	511	502	505	490
Belgium	510	503	520	501
Canada	534	524	527	527
Czech Republic	513	512	510	483
Denmark	496	485	513	494
Finland	563	556	548	547
France	495	502	496	488
Germany	516	505	504	495
Greece	473	479	459	460
Hungary	504	508	491	482
Iceland	491	470	506	484

Ireland	508	510	501	517
Italy	475	478	462	469
Japan	531	533	523	498
Korea	522	522	547	556
Luxembourg	486	483	490	479
Mexico	410	435	406	410
Netherlands	525	515	531	507
New Zealand	530	528	522	521
Norway	487	474	490	484
Poland	498	510	495	508
Portugal	474	492	466	472
Slovak Republic	488	495	492	466
Spain	488	499	480	461
Sweden	503	496	502	507
Switzerland	512	508	530	499
Turkey	424	463	424	447
United Kingdom	515	508	495	495
United States	489	483	474	n/a
Argentina <sup>Partner</sup> countries	391	416	381	374
Azerbaijan	382	388	476	353
Brazil	390	424	370	393
Bulgaria	434	446	413	402
Chile	438	465	411	442
Colombia	388	411	370	385
Croatia	493	497	467	477
Estonia	531	527	515	501
Hong Kong- China	542	560	547	536
Indonesia	393	425	391	393
Israel	454	448	442	439
Jordan	422	438	384	401

Kyrgyzstan	322	340	311	285
Latvia	490	491	486	479
Liechtenstein	522	n/a	525	510
Lithuania	488	487	486	470
Macao-China	511	523	525	492
Montenegro	412	412	399	392
Qatar	349	n/a	318	312
Romania	418	431	415	396
Russian Federation	479	483	476	440
Serbia	436	440	435	401
Slovenia	519	513	504	494
Chinese Taipei	532	546	549	496
Thailand	421	461	417	417
Tunisia	386	408	365	380
Uruguay	428	446	427	413

PISA 2007b pp175-178

PISA calculated individuals' socio-economic background as measured by its index of economic, social and cultural status (ESCS) for the 2009 survey using students' answers to the context questionnaire that asked about their parents' education and occupations, availability of certain study aids in students' homes and the presence of certain household items, used as a proxy measure for parental income or wealth. (OECD 2009b) The ESCS index was standardised to have a mean of zero for all students in the OECD and a standard deviation of one. Within the OECD national average ESCS ranged from -1.22 in Mexico to 0.47 in Norway and 0.72 in Iceland. Among partner countries it ranged from -1.55 in Indonesia to 0.42 in the UAE and 0.51 in Qatar (OECD 2009a).

PISA carried out an adjustment of national mean student reading scores but whereas for 2006 this adjustment (to science scores) was made if the mean ESCS would be equal in all countries, for 2009 a predicted reading performance for a student with a socio-economic background equal to zero (the OECD average) was calculated for each country. As can be seen in Table 3, relative to national mean reading performance, this lowered

values for countries with higher average socio-economic backgrounds and raised them for countries with lower average socio economic backgrounds.

Table 3.

Mean national scores for reading, mathematics, science and predicted reading scores for students in participating countries if they had the OECD average PISA Index of Economic, Social and Cultural Status score, PISA 2009.

	Mean Reading Score	Mean mathematics score	Mean science score	Predicted reading performance for a student with a socio-economic background equal to zero, the OECD average
Australia OECD Nations	515	514	527	502
Austria	470	496	494	468
Belgium	506	515	507	499
Canada	524	527	529	510
Chile	449	421	447	468
Czech Republic	478	493	500	483
Denmark	495	503	499	485
Estonia	501	512	528	497
Finland	536	541	554	525
France	496	497	498	505
Germany	497	513	520	493
Greece	483	466	470	484
Hungary	494	490	503	504
Iceland	500	507	496	483
Ireland	496	487	508	496
Israel	474	447	455	480
Italy	486	483	489	490
Japan	520	529	539	522

Korea	539	546	538	544
Luxembourg	472	489	484	466
Mexico	425	419	416	456
Netherlands	508	526	522	499
New Zealand	521	519	532	519
Norway	503	498	500	487
Poland	500	495	508	512
Portugal	489	487	493	499
Slovak Republic	477	497	490	482
Slovenia	483	501	512	481
Spain	481	483	488	491
Sweden	497	494	495	485
Switzerland	501	534	517	498
Turkey	464	445	454	499
United Kingdom	494	492	514	488
United States	500	487	502	493
Albania <sup>Partner</sup> countries	385	377	391	416
Argentina	398	388	401	424
Azerbaijan	362	431	373	376
Brazil	412	386	405	445
Bulgaria	429	428	439	437
Colombia	413	381	402	445
Croatia	476	460	486	482
Dubai (UAE)	459	453	466	439
Hong Kong- China	533	555	549	548
Indonesia	402	371	383	428
Jordan	405	387	415	420
Kazakhstan	390	405	400	410

Kyrgyzstan	314	331	330	341
Latvia	484	482	494	488
Liechtenstein	499	536	520	497
Lithuania	468	477	491	471
Macao- China	487	525	511	495
Montenegro	408	403	401	416
Panama	371	360	376	402
Peru	370	365	369	424
Qatar	372	368	379	360
Romania	424	427	428	437
Russian Federation	459	468	478	468
Serbia	442	442	443	440
Shanghai- China	556	600	575	569
Singapore	526	562	542	547
Chinese Taipei	495	543	520	507
Thailand	421	419	425	450
Trinidad & Tobago	416	414	410	441
Tunisia	404	371	401	426
Uruguay	426	427	427	453

(OECD 2010c Annex B1, 2010a)

Some basic information such as school size or proportion of females enrolled is contained in enrolment data submitted by the school. Other information sought by PISA is obtained through a questionnaire completed by principals. Information sought includes:

- Admittance policies: the degree to which factors such as residence in the local area, academic record and parents' endorsement of the school's religious philosophy of the school determine students' admittance to the school
- School type: public, private, or government-dependent private.

-Autonomy: Whether decisions about resources and curriculum were made by principals and teachers, a school board or regional or national authorities.

-School Choice: whether the school competes with others in the local area for students, and if so the number.

-Accountability Arrangements: whether achievement data were tracked over time by an administrative authority, whether such data were used in the evaluation of the teachers' or principal's performance, whether the data was posted publicly and whether such data were used in decisions about instructional resource allocation to and within the school. (OECD, 2007b)

For PISA 2006, on average 300 principals were surveyed in each OECD country and in seven countries fewer than 170 were surveyed. For PISA 2009 on average 264 principals were surveyed and in six countries fewer than 150 were surveyed. Data derived from school principals and students were weighted to reflect the number of 15-year-olds in each school. (OECD, 2007b, OECD 2010b)

#### **4.9. Discussion of PISA methodology**

PISA is confident that its methods are sound and its measures of student performance are valid, and is strident in defence of these (Adams, 2003). At the same time, PISA advises a general caution in interpreting its results as many important contextual factors cannot be captured by surveys of this kind and thus cause and effect cannot be firmly established. The learning environment of a 15 year old at the time they undertake PISA is not necessarily the same as that which shaped their learning earlier in their education and as such the contextual data collected is an "imperfect proxy for the cumulative learning environments of students" (OECD, 2007b p.214-15). Smith (2006), while using PISA data for secondary analysis, sounds a general caution on the limits of cross national education surveys such as PISA, as

The virtual absence of cross national educational variables that predict student achievement suggest it is difficult to separate the effect of educational policies and instructional practices from the contexts in which they are developed and implemented (p.36)

PISA also warns of the need to take into account the limitations of the contextual data gathered through principal surveys. This is because on average only 300 principals were surveyed in each country in 2006 and 264 in 2009, and the data gathering relies on a principal's assessment of school resource matters such as class size, for which they may not have accurate or up-to-date data (OECD 2007b). Caution in regard to the results of the principal survey should surely extend further to note that they are based on principals' subjective responses to being questioned whether they had 'considerable responsibility' for determining such things as course content. White and Smith (2005) are conscious of this limitation on data obtained from the principal survey but still believe it provides the best quality data available for their area of investigation, teacher supply and shortage.

PISA recommends the results concerning the effect of school choice on performance be treated with caution as the existence of other schools within a local area does not necessarily mean that parents have access to these, particularly if they are private schools. (OECD, 2007b)

Also, PISA acknowledges it is necessary to take into account the lower retention rates of 15 year old students in less developed, predominantly non-OECD countries (OECD 2009). PISA raises this point as a possible limitation on the comparability of its findings, particularly when comparing OECD and less developed countries, but does not discuss the matter in detail. For Wuttke (2007) and Prais (2003) this is a serious shortcoming. Prais calls PISA's setting of the target population age at 15 years and 3 months to 16 years and two months an error of judgement as non-retention in countries including Brazil and Mexico at this age is close to 50 per cent and among this age group, other countries, including Britain, experience high levels of absenteeism. Students who have left school at this age and have no chance of taking part in PISA are likely to be academically weaker, challenging PISA's claim to be measuring the performance of education systems near the end of compulsory schooling. The choice of the target age population is also questioned on the grounds that at 15 years old, the ability of abstract reasoning is still developing, and is hence underestimated (Wuttke, 2007) and that mathematics education in European countries such as Germany is likely to be more conceptual at this age compared with Britain where maths is compulsory only to age 16 and likely to be taught to 15 year olds with greater emphasis on real-life applications of the kind PISA assesses (Prais 2003).

Using age, rather than schooling grade, to define the target population is criticised on the grounds it does not acknowledge those children could be spread across up to three grades and at very different stages of learning and development (Prais, 2003). Adams (2003) defends this choice on the grounds that cross country differences in pre-school provision and school starting ages mean different grades would have to be selected in different countries, creating a greater comparability problem.

In discussing variation in PISA 2003 student response rates of up to 15 per cent across countries, Wuttke (2007) writes that ‘non response bias can be considerable because the propensity of school principals and students to partake in testing is likely to be correlated with the potential outcome’ (p.9). Prais (2003, 2004) agrees with this likely non-response bias and raises concerns with PISA’S method of calculating of its response rate in which replacement schools taking part are included in the calculation of a response rate based on the number of schools that ended up participating as a proportion of the initial school sample size. For Britain in 2000, the initial school response rate was 61 per cent. 55 per cent of replacement schools approached then also took part. PISA gave its response rate as 80 per cent while Prais believes it should be defined as 57 per cent. Prais (2003) is also critical of PISA’S exclusion of schools with 25 per cent – 50 per cent participation from school level calculations while the students’ individual results were used in the calculation of mean national scores. Adams (2003) acknowledges the response rates of Britain, whether calculated before or after replacement, were of concern but claims the direction and magnitude of non-response bias cannot be known and analyzes the academic attainment of non-responding British government schools, based on GCSE attainment levels, finding no evidence they were on average lower achieving. PISA 2003 response rates in Britain and the USA were, according to Smith (2006), too low for findings to generalised to the larger population.

Response rates continue to be an issue for PISA. In the 2009 survey school response rates after replacement were below the required level of 85% in the United States and Panama and the student response rate was below the PISA-determined 80% requirement in Canada. Despite this, and after what it terms additional analysis, PISA retained results from these countries in the final database and used them in its analysis (OECD 2011).

The PISA survey’s reliance on item response theory – the assumption that the probability of a correct response depends only on the differences in students’ competence and the

item's degree of difficulty – is criticised by Wuttke (2007) for insufficient consideration of the impact of non-competence measures such as items being easier for one subpopulation than another. The reason for its use in PISA is 'neither theoretic or empiric but pragmatic: only one dimensional models yield unambiguous rankings' (p.11). The design of the PISA survey and the treatment of results, oriented as they are toward the creation of mean national scores and country rankings, are questioned. For Goldstein (2004) international comparative surveys provide more insight when they facilitate the investigation of underlying differences between countries, something that requires a more complex, multidimensional approach than that of PISA. The statistical construction of PISA's scales to have a mean of 500 and an standard deviation of 100, with the difficulty of questions being determined by the proportion of students getting an answer correct, means they are relative, not absolute, measures of achievement. Dohn (2007) criticises PISA's failure to verify that statistical levels on its scales of difficulty correspond to actual ones. Prais (2003, 2004) calls for PISA to publish country figures on the correct average scores for each question for a clearer view of countries' educational weaknesses and strengths and to make PISA's scaling technique more transparent. Wuttke (2007) calls PISA's proficiency levels 'arbitrary' and creation of its scales 'ill documented renormalisation' (p.17).

The PISA survey is accused widely of containing cultural and linguistic bias (Dohn 2007, Goldstein 2004, Hopmann and Brinek 2007, Prais 2003, Wuttke 2007). Other criticisms include gender imbalances in PISA 2003, when only 40.5% of participants in South Korea were girls and subjective judgement of, and widely varying levels of, exclusions due to intellectual impairment (Wuttke 2007).

Hopman and Brinek (2007) write that while the establishment of PISA is an important milestone in comparative education "the product of most public value, the national league tables, are based on so many weak links they should be abandoned" and any policy making based on them cannot be justified (p.13). This viewpoint is representative of what can called be called PISA sceptics (Dohn 2007, Prais, 2003 & 2004, Wuttke 2007) while at the other end of the spectrum, PISA is dismissive of much of the criticism (Adams, 2003) and while it does discuss limitations of its own methodology in the Analysis volume, findings such as those that endorse various aspects of school autonomy as being associated with higher student performance, are stated without reservation. In between could be said to be researchers like Goldstein (2004) and Smith (2006) who recognise

PISA's limitations but still believe it generates valid data to help understanding of different school systems and facilitate secondary analysis.

## **5. 2006**

### **5.1 PISA's Analysis and Results**

The country rankings, or league tables, based on national mean student scores that dominate media coverage when PISA findings are released are really only the tip of the PISA iceberg. The Analysis volumes that accompany the release of results contain detailed descriptions of student performance in the context of a wide variety of demographic, socio-economic and school organisation data.

Science was the focus of PISA 2006 and a more detailed profile of students' performance in this subject area was developed. As well as the combined science scale, scores on which were used to determine the national mean science score, different scales assessing particular science competencies were employed. Science attitudinal questions enabled PISA to develop a detailed profile of student engagement in science including how much 15 year olds value science, believe they can succeed in science and feel responsible towards resources and the environment. This was done using descriptive statistics; for example, giving national percentages of students who agreed or strongly agreed with the statement "I enjoy acquiring knowledge in science." Scatterplots were also used to illustrate the relationship between selected attitudes towards science and the environment and performance in science. (OECD 2007b)

Using its 2006 data, PISA conducted analyses of relationships between school and system characteristics and student performance in science, reading and mathematics but states that since the results did not vary fundamentally across the different subject areas, the Analysis volume only discusses the relationships with science performance (OECD 2007b).

Between-school and within-school variance in science performance in participating countries, expressed as a percentage of the average variance in student performance in OECD countries, were calculated as a measure of equity. Deeper investigation by PISA of factors related to variation in performance revealed some interesting results. For example, comparison of mean science scores for native and first generation immigrant students showed the latter lag, on average, 58 points behind (OECD, 2007b).

In looking at how socio-economic background relates to student performance, PISA used several analytical statistic methods. When science performance was graphed against socio-economic background for all students in the OECD, an increase of one standard deviation in the PISA ESCS index was, on average, associated with a 40 points higher science score (OECD, 2007b).

### **5.1.1 Science Performance and School and System Characteristics**

As part of its study of ‘school and system characteristics and student performance in science’, PISA collected data on the implementation of many school organisation matters including resources (i.e. class sizes) and grouping policies and investigated the relationships of these with student results on the PISA test. Most of its interest in the area of school organisation though is in matters central to the GERM. Included as ‘Key Findings’ in PISA 2006’s Executive Summary were the following observed relationships between school organisation policies and science performance:

- ‘Across countries, having a larger number of schools that compete for students is associated with better results, over and above the relationship with student background’
- After other factors are taken into account ‘there still remains a significant positive association between schools making their achievement data public and having stronger results.’
- ‘Students in countries where autonomy is more common tend to do better in the science assessment, regardless of whether or not they themselves are enrolled in relatively autonomous schools’ (OECD 2007a, pp. 7-8)

### **5.1.2 School Competition**

PISA uses data aggregated at the individual, school and system/country level in its analysis. In its 2006 analysis of school competition effects, PISA determined a school with a high level of competition as one competing with one or more other schools for students, based on responses to the principal survey. Levels of school choice, defined in this way, varied from 95 per cent of students in Indonesia attending schools where the

principal reported competing with one or more other schools for students in the same area, to 34 per cent in Norway. The OECD average was 73 per cent (OECD, 2007b).

Through a comparison of means, at the school level, PISA found students at schools with a high level of competition performed on average 17.9 points better on the science scale. This difference however was found to not be statistically significant after the following demographic and socio-economic background factors are accounted for: students' ESCS, gender, students' and parents country of birth, and the language spoken at home; at the school level, the socio-economic intake of the school, the school location and the school size; and at the country level, the national average ESCS. In analysing school competition at the national level, PISA looked at the change in national mean science scores per countries having an additional 10% of competing schools. After accounting for demographic and social factors, this difference was found to be 6.7 points and statistically significant (OECD, 2007b). (Note: when analysing the effect of system level factors using this method, PISA considered results statistically significant at the country level if the p-value was below 0.1) This result is the basis of the PISA 2006 finding on school competition.

### **5.1.3 School accountability policies**

Posting school data publicly was one of nine accountability policies for which PISA investigated the relationship with science performance. Based on answers to the principals' survey, schools were classified as posting achievement data publicly or not. The change in score for schools posting was 14.7. After accounting for demographic and social factors this difference was 6.6 and remained statistically significant (OECD, 2007b). The same analysis method was used to investigate the relationships between science results and whether schools implemented a variety of accountability policies. These were

- School informing parents of children's performance relative to other students in the school
- School informing parents of students' performance relative to other schools
- School informing parents of students' performance relative to national benchmarks

- School using achievement data for evaluating principals
- School using achievement data for evaluating teachers
- School using achievement data for allocating resources
- School with achievement data tracked over time

For all these variables no significant relationship was observed.

#### **5.1.4 School autonomy**

The principals' survey asked whether schools had considerable responsibility, both schools and regional and/or national education authorities had considerable responsibilities, or only regional and/or national education authorities had considerable responsibilities for these aspects of school management:

- Selecting teachers for hire
- Dismissing teachers
- Establishing teachers' starting salaries
- Determining teachers' salary increases
- Formulating the school budget
- Deciding on budget allocations within the school
- Establishing student disciplinary policies
- Establishing student assessment policies
- Approving students for admission to the school
- Choosing which textbooks are used
- Determining course content
- Deciding which courses are offered

Cross-country correlations were calculated between the percentage of schools in countries having considerable responsibility ('school only' and 'school and government') for each of the above and student performance in science. As Table 4 shows, the correlations were positive and statistically significant in the case of every variable except establishing and determining teachers' salaries. This predominance of statistically significant positive correlations between these various aspects of school autonomy and science performance is the basis of the finding that "the data suggest that in those countries in which principals

reported, on average, higher degrees of autonomy...the average performance in science tends to be higher.” (OECD 2007b, p.249)

Table 4.

Correlations between degree of countries’ adoption of selected school autonomy measures and student science performance, PISA 2006

<b>Aspect of school management</b>	<b>Cross-country correlation between the percentage of schools having considerable responsibility (‘school only’ and ‘school and government’) and science performance</b>
Formulating the school budget	0.47*
Dismissing teachers	0.32*
Selecting teachers for hire	0.43*
Establishing teachers starting salaries	0.20
Determining teachers’ salary increases	0.22
Deciding on budget allocations within the school	0.54*
Establishing student disciplinary policies	0.41*
Establishing student assessment policies	0.43*
Approving students for admission to the school	0.27*
Choosing which textbooks are used	0.51*
Determining course content	0.52*

Deciding which courses are offered	0.58*
------------------------------------	-------

\*correlation coefficient is statistically significant at the 5% level ( $p < 0.05$ ) (OECD, 2007c Table 5.10 pp175-178)

## 5.2 Secondary analysis using PISA 2006 data

### 5.2.1 Development of research design

As discussed above, PISA uses student, school and system/country level data and key findings are based on analysis of data collected at all of these levels. Examining the PISA 2006 Data Volume and the PISA website I realised there was data available on the degree to which countries implemented a number of GERM policies. These are mostly expressed as the percentage of students within a country studying at schools which operate under a given policy, based on school principals' answers to survey questions. This availability enabled the calculation of cross-country correlations between the degree of implementation of selected GERM policies (independent variables) and performance in science, reading and mathematics (dependent variables).

In 2006 PISA did not conduct correlational analysis of the relationships between school organisation policies other than autonomy (measured as a percentage of schools within a country having responsibility for various things) and performance, or if it did, it did not publish the results. In fact, PISA looked at school autonomy's relationship with science performance at both the school and system/nation level. Among schools it looked at the effect of a one standard deviation increase in indices of school autonomy it derived (staffing, educational content, budgetary and a combined index), on student performance. For all four measures of autonomy no significant relationship was observed. However, as shown above, for a range of autonomy variables, PISA also conducted cross country correlations which led to its finding that 'in countries where autonomy is more common (students) tend to do better in the science assessment, regardless of whether or not they themselves are enrolled in relatively autonomous schools' (OECD 2007a p.42, 2007b).

PISA did not carry out this dual level of analysis (school and national system) for other school organisation characteristics such as the accountability measures it collects data on. Through my investigation of the PISA 2006 data set I realised that data was available that

would enable me to investigate relationships at the system/national level between student performance and a representative range of GERM-type policies:

- Percentage of children in schools posting achievement data publicly.
- Percentage of children in schools where the governing board exerts direct influence over decision making about staffing. This is a measure of local autonomy and independence from centralised control. Other measures of autonomy are not included as PISA published cross-country correlations between these and student performance.
- Percentage of children in private schools.
- Percentage of schools competing for students with two or more other schools in the same area.
- Percentage of students in schools where the principal reported that achievement data are being used in decisions about instructional resource allocation to the school. This a measure of the prevalence of consequential accountability measures targeted at schools.
- Percentage of students in schools where the principal reported that achievement data are being used for evaluation of teachers' performance. This is a measure of the prevalence of consequential accountability measures targeting teachers, including performance pay.

The general hypothesis to be tested was that greater implementation of a given GERM school organisation policy is associated with the higher student performance as measured by PISA.

I would analyse the relationship between these policies and student performance in all three subject domains and the PISA derived mean science scores if the mean ESCS would be equal in all OECD countries.

Correlations would be tested for by calculating the Pearson correlation coefficient ( $r$ ) between the levels of implementation of the selected GERM policies and student performance. A significance level of 0.05 is used, the same level PISA used to test the significance of its cross country correlations, although it is also reported where results are significant at the  $p < .01$  level.

### 5.2.2 Independent variables

- Percentage of children in schools posting results publicly. This ranged from 4.5 % in Finland to 92.7 % in the United Kingdom. The average for OECD countries taking part was 38.1 %. For this calculation  $N=55$  as data on result posting not available for France or Liechtenstein (OECD 2007c, Table 5.8, p. 173).
- Percentage of children in schools where the governing board exerts direct influence over decision making about staffing. This ranged from 0.6% in Poland to 90.6% in Hungary. The average for OECD countries was 33.8 %. For this calculation  $N=55$  as data on governing board influence not available for France or Liechtenstein (OECD 2007c, Table 5.12a, p.180).
- Percentage of children in private schools. The figure used was the combined total percentage of students in government-dependent private schools and government-independent private schools in each participating country. This ranged from 0 in Latvia, Romania and Russia to 96.2 % in Macao-China. The average for OECD countries was 14.4%. For this calculation  $N=52$  as data on students in private schools was not available for Australia, Belgium, France Bulgaria or Liechtenstein (OECD 2007c, Table 5.4, p165-166).
- Percentage of schools competing for students with two or more other schools in the same area. This ranged from 21.8% in Norway to 90% in Indonesia. The average for OECD countries was 60.3%. For this calculation  $N=55$  as data on school competition was not available for France or Liechtenstein. (OECD 2007c, Table 5.5, p.168).
- Percentage of students in schools where the principal reported that achievement data are being used in decisions about instructional resource allocation to the school. This ranged from 1.5% in Greece to 86.5% in Chile. The average for

OECD countries was 30.2%. For this calculation N=55 as data not available for France or Liechtenstein (OECD 2007c, Table 5.8, p.173).

- Percentage of students in schools where the principal reported that achievement data are being used for evaluation of teachers' performance. This ranged from 4.8% in Luxembourg to 100% in Serbia. The average for OECD countries was 43.3%. For this calculation N=55 as data was not available for France or Liechtenstein (OECD 2007c, Table 5.8, p.173).

### **5.2.3 Dependent variables**

- Mean national science score
- Mean national reading score
- Mean national mathematics score
- PISA-calculated hypothetical science score if the PISA index of economic, social and cultural status were equal across all OECD countries.

### **5.2.4 Results of secondary analysis using 2006 data**

Results of correlations can be seen in Table 5.

The significant positive cross country correlations between performance and percentage of students in schools competing with two or more other schools were consistent with PISA 2006's finding on the effect of school competition. PISA found that across countries, having an additional 10 per cent of schools competing with one or more schools was associated with, on average, a science score higher by 6.7 points and that this difference was significant once socio-economic and demographic factors were taken into account.

Cross-country correlations between the percentage of students in schools posting results publicly and student performance were not significant.

Cross-country correlations between the percentage of students in schools in which the governing board exert direct influence over decision making about staffing and student performance were not significant.

The cross country correlation between the percentage of students in private schools and student performance was not significant for science, reading and mathematics scores. The correlation between numbers in private schools and science score if the mean ESCS would be equal across all OECD countries was positive and statistically significant.

Table 5.

Correlations between level of implementation of GERM policies in participating countries and student performance, PISA 2006, all 57 participating countries

		Science performance	Reading performance	Maths Performance	Science Performance if ESCS was equal across all OECD countries
Percentage of students in schools where:	Student Results posted publicly	-0.09	-0.14	-0.06	-0.12
	Government Board exert influence over staffing decisions	0.21	0.16	0.21	0.26
	School competing	<b>0.32*</b>	<b>0.31*</b>	<b>0.31*</b>	<b>0.33*</b>
	Achievement data used for teacher evaluation	<b>-0.54**</b>	<b>-0.56**</b>	<b>-0.52**</b>	<b>-0.47**</b>
	Achievement data used for resource allocation	<b>-0.41**</b>	<b>-0.37**</b>	<b>-0.40**</b>	<b>-0.37**</b>
Percentage of students in private schools		0.21	0.25	0.23	<b>0.29*</b>

\*Correlation is significant at the 0.05 level

\*\*Correlation is significant at the 0.01 level

The cross-country correlations between student performance and the percentage of students in schools using student data for teacher evaluation and resource allocation were negative and significant in all cases. The strong negative correlations between the degrees of implementation of these policies and all measures of student performance are statistically significant at the 1 per cent level ( $p < .01$ ).

### **5.3 Summary**

PISA 2006 employed a variety of statistical methods and used data aggregated at different levels and different significance tests to arrive at its key findings on relationships between school and system characteristics and student performance.

PISA arrived at its key finding that school autonomy is associated with better student performance by performing cross-country correlation analysis alongside school level analysis of the effect of a one-standard deviation increase in measures of autonomy on student performance. By adapting the cross-country correlational analysis that PISA carried out, it was possible to look at the relationship between a range of school organization policies favoured by the GERM and student performance. In effect this expanded the dual level of analysis (school and national/system) PISA conducted of school autonomy, to a wider variety of school policies.

The significant positive cross-country correlation observed between school competition and all measures of student performance is consistent with PISA's school level analysis of the effect of competition.

The strong and statistically significant negative correlations between the two consequential accountability measures – the use of student achievement data to allocate school resources and to evaluate teachers – and student performance are not consistent with PISA's analysis of the effect of these policies at the school level. This inconsistency also was observed in PISA's analysis of autonomy measures, yet the significant positive cross-country correlations between autonomy and student performance were the basis for a 'key finding'.

## 6. 2009

### 6.1 PISA's Analysis and Results

Reading was the main focus of the 2009 survey, with science and mathematics also tested. Again, students and school principals completed the questionnaires that provided the contextual information used by PISA in its analysis. Data collected through these has been made public and used in the current study (OECD 2010b).

In its analysis of relationships between school organisation policies and learning outcomes, PISA made greater use of cross country correlations with the 2009 data. In doing so it carried out a number of analyses using the 2009 data that I had first carried out using the 2006 data and presented as part of a master's thesis seminar at the University of Oulu in November 2008. The PISA 2009 Table IV.2.1 *Correlations between system-level characteristics and educational outcomes* can be considered an expanded version of my Table 5 above, used to present results of correlations using PISA 2006 data.

In making broad use of the method I had employed to investigate relationships between school organisation policies and student performance, and discontinuing many of its more detailed statistical analyses, PISA expresses its confidence in the 'robustness and sensitivity' of its findings and states that 2009 system-level correlations were found to be consistent with 2006 results achieved using more sophisticated statistical techniques (OECD 2010b p.30).

PISA compiled selected results of 2009 correlations in Table IV.2.1 *Correlations between system-level characteristics and educational outcomes* (OECD 2010b Annex B1). As can be seen from Table 6 this included the correlations between reading performance and countries' average index of school responsibility for curriculum and assessment and average index of school responsibility for resource allocation (As results across the three domain areas were very similar, PISA only published the results of analysis using reading performance, the 2009 focus area).

Table 6.

Selected results from PISA 2009 Table IV.2.1, *Correlations between system level characteristics and educational outcomes*

		Reading Performance- All countries	Reading Performance - OECD Countries	
School governance	Average index of school responsibility for curriculum and assessment	<b>0.46</b>	<b>0.45</b>	
	Average index of school responsibility for resource allocation	0.25	0.02	
	Percentage of students in private schools	0.14	0.05	
	Percentage of students in schools that compete with other schools in the same area	0.16	0.06	
Assessment and accountability policies	Percentage of students in schools that assess students with standardised tests	-0.05	0.15	
	Percentage of students in schools that use assessment data to:	Compare the school with other schools	-0.21	0.02
		Allocate resources	<b>-0.47</b>	-0.08
		Monitor teacher practices	<b>-0.46</b>	-0.17
		Post achievement data publicly	-0.11	0.04
		Monitor progress over time	<b>-0.29</b>	-0.09
		Have their progress tracked by administrative authorities	<b>-0.36</b>	-0.14

Results in bold are significant at the 5% level ( $p < 0.05$ ), results in italic are significant at the 10% level ( $p < 0.10$ )

(OECD 2010b Annex B1)

The index of school responsibility for curriculum and assessment is a composite measure of schools' autonomy over establishing student assessment policies, choosing textbooks, determining which courses are offered and the content of those courses. The index of school responsibility for resource allocation is a composite measure of schools' autonomy

over appointing and dismissing teachers, establishing teachers' starting salaries and salary raises, formulating school budgets and allocating them within the school. These indices are created to have a mean of zero and standard deviation of one for all OECD countries and are calculated for each participating school based on principals' questionnaire answers. From these, PISA calculated national average scores on the indices and used these in its cross country correlations with student reading performance.

PISA calculated the correlation coefficient between reading performance and average index of school responsibility for curriculum and assessment was 0.46 across all countries and 0.45 across OECD countries and in both cases was statistically significant at the 5% level ( $p < 0.05$ ). This, PISA found, was a 'clear relationship' that enabled it to find that "In countries where schools have greater autonomy over what is taught and how students are assessed, students tend to perform better" (OECD 2009b p.14). No significant relationship was observed between greater autonomy to allocate resources and student performance. (OECD 2009b)

In its expanded use of cross country correlations, PISA looked at the relationship between student reading performance and a variety of school organisation policies, many of them part of the GERM. No significant relationship was observed between reading performance and the following school system characteristics:

- The percentage of students in schools competing with other schools in the same area
- The percentage of students in private schools
- The percentage of students in schools that assess students with standardised tests
- The percentage of students in schools posting achievement data publicly.

Significant ( $p < 0.05$ ) negative correlations were observed between reading performance and the percentage of students in schools that use student assessment and achievement data for a variety of accountability purposes:

- Monitor progress over time ( $r = -0.29$ )
- Have their progress tracked by administrative authorities ( $r = -0.36$ )
- Allocate resources ( $r = -0.47$ )
- Monitor teacher practices ( $r = -0.46$ )

When calculated for OECD countries only, these relationships were not significant (OECD 2010b Annex B1).

## **6.2 Secondary analysis using PISA 2009 data**

Data on the influence school boards have over matters including staffing was not available for PISA 2009. This is part of a rearrangement of how data is presented. For PISA 2006, a large amount of contextual data was presented plainly in Volume 2 of the PISA report. For 2009, a more limited range of contextual data was made available across six volumes which also contain discussion and analysis. Hence it was not possible to repeat the correlation between the percentage of children in schools where the governing board exerts direct influence over decision making about staffing and student performance. It is however possible to investigate relationships between student performance and the other school organisation policies used in analysis of the 2006 results using the same method.

As part of its adoption of the cross country correlation method of analysis in 2009, PISA calculated some of the same correlations that I carried out allowing for verification of my analysis. In all cases where the same correlation was calculated, the results were identical.

### **6.2.1 Independent Variables**

- Percentage of children in schools posting results publicly. This ranged from 0.6% of students in Shanghai-China, 1.9% in Belgium and 2.5% in Finland to 89.3% in the United States and 86% in Azerbaijan. The average for OECD countries was 36.6%. For correlations using this variable N=64 as data not available for France (OECD 2010b Annex B1 Table IV 3.13).
- Percentage of children in private schools. The figure used was the combined total percentage of students in government-dependent private schools and government-independent private schools in each participating country. This ranges from 0.1% in Russia and 0.4% in Azerbaijan and Lithuania to 96% in Macao and 92.5% in Hong Kong. For correlations using this variable N=63 as data not available for Belgium and France (OECD 2010b Annex B1 IV.3. 9).

- Percentage of students in schools where principal reported competing for students with two or more other schools in the same area. This ranged from 10.2% in Liechtenstein and 22.3% in Norway to 90.2% in Australia, 90.3% in Singapore and 94% in Hong Kong. The OECD average was 61.2%. For correlations using this variable N=64 as data not available for France (OECD 2010b Annex B1 Table IV 3.8a).
- Percentage of students in schools where the principal reported that achievement data are being used in decisions about instructional resource allocation to the school. This ranged from 1.2% in Iceland and 3.5% in Greece to 90.9% in Indonesia and 89.5% in Singapore. The OECD average was 32.7%. For correlations using this variable N=64 as data not available for France (OECD 2010b Annex B1 Table IV 3.13).
- Percentage of students in schools where the principal reported that achievement data are being used for evaluation of teachers' performance. This ranged from 0 in Liechtenstein and 8.2% in Luxembourg to 97.5% in Kazakhstan and 97.4 in Albania. The average for OECD countries was 44.8%. For correlations using this variable N=64 as data not available for France (OECD 2010b Annex B1 Table IV 3.13).

### **6.2.2 Dependent variables**

- Mean national science score
- Mean national reading scores
- Mean national mathematics score
- PISA-calculated predicted reading score for a student with a PISA index of economic, social and cultural status background equal to zero

### **6.2.3 Results of secondary analysis using 2009 data**

2009 results are collated in Table 7.

Table 7.

Correlations between level of implementation of GERM policies in participating countries and student performance, PISA 2009, all participating countries

		Science performance	Reading performance	Maths Performance	Predicted reading score for a student with ESCS= 0
Percentage of students in schools where:	Student Results posted publicly	-0.11	-0.11	-0.08	-0.18
	School competing	0.16	0.17	0.14	0.23
	Achievement data used for teacher evaluation	<b>-0.49**</b>	<b>-0.50**</b>	<b>-0.49**</b>	<b>-0.46**</b>
	Achievement data used for resource allocation	<b>-0.48**</b>	<b>-0.47**</b>	<b>-0.48**</b>	<b>-0.41**</b>
Percentage of students in private schools		0.14	0.14	0.14	0.16

\*Correlation is significant at the 0.05 level \*\*Correlation is significant at the 0.01 level

Cross-country correlations between the percentage of students in schools posting results publicly and all measures of student performance were not significant.

Cross-country correlations between percentage of students in schools competing with two or more other schools and all measures of student performance were not significant.

Cross-country correlations between the percentage of students in private schools and all measures of student performance were not significant.

The cross-country correlations between all measures of student performance and the percentage of students in schools using student data for teacher evaluation and resource allocation were negative and significant in all cases. The strong negative correlations between the degrees of implementation of these policies and all measures of student performance are statistically significant at the 1 per cent level ( $p < .01$ ).

### **6.3 Summary**

PISA's adoption of the system level/cross-country correlation method to investigate relationships between student performance and a wide range of school organisation policies made its 2009 analysis more systematic and consistent than in 2006. As well as providing support for my decision to use this method beginning with with 2006 data, PISA's use of it enabled the checking of my results against theirs. In all cases where the same correlation was calculated, the results were identical.

Using 2009 data, PISA analyses the same relationships as I do but with one important exception. In looking at what it terms assessment and accountability policies, PISA did not conduct the correlation between percentage of students in schools that use student data for teacher evaluation and reading performance. This is surprising as PISA collected the data to enable this analysis to be done and it could have been added to the list of correlations involving accountability policies in Table IV.2.1. This table included a correlation between reading performance and the use of student achievement data to monitor teacher practices, a policy which was not mentioned in PISA's 2006 Analysis and which does not attract the same interest from researchers and educators as use of student data for evaluating teachers. As I have shown, this correlates negatively with student performance and this relationship is significant. The same correlation carried out across OECD countries for all subject domains is negative and significant ( $r = -.39$ ) in the case of mathematics performance (see Appendix 1).

PISA's finding of a 'clear relationship' between school responsibility for curriculum and assessment and student performance is based on a significant positive correlation with reading scores of 0.46 across all countries. Such autonomy is thus in PISA's (2010b) words 'a feature shared by high performing schools' (p.27). This finding is supported by a significant positive correlation across OECD countries.

The significant negative correlations across all countries that PISA found between the use of student achievement data to allocate resources, monitor progress over time and monitor teacher practices are not presented as findings. Remarkably, these significant negative correlations are not even mentioned in PISA's analysis volume (OECD 2010b).

## 7. CONCLUSIONS

This secondary analysis of data from PISA's 2006 and 2009 surveys looked at the relationship between selected school organisation policies advocated by the global education reform movement and student performance, producing mixed results.

Posting student results publicly and the level of private schooling within a country do not show a significant relationship with student performance. Increased school competition was shown to have a significant positive relationship with performance in 2006 but not 2009.

Of most interest are the findings relating to consequential accountability policies. For 2006 and 2009, the proportion of students in schools using student achievement data for teacher evaluation and for resource allocation to schools showed significant negative correlations with every measure of student performance. Bearing in mind that the current study and PISA are non-experimental research, and the limitations PISA admits are inherent in its data, these results raise doubts about the effect of these policies on student learning. Research used in debates over the merits of these policies comes largely from studies conducted at local and national levels. These results of secondary analysis of data from the largest international education survey present a challenge to advocates of the GERM who commonly claim that penalising and rewarding schools and teachers based on the results of standard tests will improve learning outcomes.

This secondary analysis was carried out using PISA's own data, in full understanding of its strengths and limitations. The decision to use cross-country correlational analysis on 2006 PISA data was vindicated by PISA's adoption of this method in its 2009 analysis. This also allowed verification of a number of correlations conducted on the 2009 data.

PISA's selective use of statistical methods in 2006 raises the question of why it performed cross-country correlations using measures of school autonomy but not other school organisation policies for which it had collected data. This situation was remedied somewhat in 2009 by the decision to calculate correlations between reading performance and a broad range of school organisation policies in a more systematic fashion. Remarkably though, PISA did not calculate the correlation between the percentage of schools in a country using performance data for teacher evaluation and reading

performance (or if it did carry out this analysis, it did not publish the result). This occurred despite the fact PISA had collected the data to enable this analysis. This accountability measure is one of the most contentious schooling policies being advanced today and the subject of much scholarly debate. There is no excuse for not including it in the correlations. As mentioned above, secondary analysis of PISA 2009 data showed use of student achievement data for teacher evaluation correlated negatively with student learning outcomes and the relationship was significant.

PISA is not consistent in the manner in which it treats significant correlations emerging from the 2009 data. The significant positive relationship between reading performance and educational and curricular school autonomy is publicised in the executive summary. The significant negative relationships between reading performance and use of student achievement data to allocate resources, monitor progress over time and monitor teaching practices appear once in a table in an annex and are never mentioned again.

The treatment of non-significant results also deserves attention. In PISA's analysis many school system characteristics show no significant relationship with performance but this does not necessarily mean the results should not be published and discussed in the Analysis volumes. School competition is an issue of debate in many countries and a policy objective of the World Bank. The fact that PISA 2009 found it has no impact on student performance is relevant and should be clearly stated.

The secondary analysis conducted for the current study, coupled with scrutiny of how PISA uses its own data and treats its findings, raise serious questions. The shortcomings in its use and reporting of its data and analysis leave PISA open to criticism that is not operating to the highest standards of objectivity and academic rigour. Repeatedly PISA has failed to conduct or publish analysis of its data that casts GERM policies in a negative light, while drawing attention to results that favour particular GERM policies, such as greater school autonomy. As such, PISA cannot be said to conduct its analysis in a meticulous manner and leaves itself open to the criticism that it is not a neutral source of policy advice but part of the global education reform movement.

## 8 . REFERENCES

- Adams, R. J. (2003). Response to 'Cautions on OECD's recent educational survey (PISA)'. *Oxford Review of Education*, 29, 377-379.
- Amrein, A., & Berliner, D. (2002). High-stakes testing, uncertainty and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved 6/2/2012 from <http://epaa.asu.edu/epaa/v10n18/>.
- Apple, M. (2004). Creating difference: Neo-liberalism, neo-conservatism and the politics of education reform. *Educational policy*, 18(12), 12-43.
- Ball, S. (1998). Big policies/small world: An introduction to international perspectives in education policy. *Comparative Education*, 34(2), 119-130.
- Ball, S. (2006). Performativities and Fabrications in the education economy: towards the performative society. In Lauder, H., Brown, P., Dillabough, J. & Halsey, A. (Eds.) *Education, globalization and social change* (pp. 692-701). Oxford: Oxford University Press.
- Benavot, A., & Tanner, E. (2007). *The growth of national learning assessments in the world, 1995-2006*. Paper for UNESCO. Retrieved 6/2/2012 from [http://campus.ph.fhnw.ch/pub/Main/UniBaseIHS09/Benavot\\_Tanner\\_2008\\_growth\\_assessments.pdf](http://campus.ph.fhnw.ch/pub/Main/UniBaseIHS09/Benavot_Tanner_2008_growth_assessments.pdf)
- Bradley, S. & Taylor, J. (2002) The effect of the quasi-market on the efficiency-equity trade-off in the secondary school sector. *Bulletin of Economic Research*, 54(3) 295-314.
- Bryman, A. (2008). The End of the Paradigm Wars. In Alasuutari, P., Bickman, L. & Brannen, J. (Eds.) *The sage handbook of social research methods*. (pp. 13-25). London:Sage.

- Burbules, N. and Torres, C., (2000). Globalization and Education: An Introduction. In Burbules, N. and Torres, C. (Eds.) *Globalization and education: critical perspectives*. (pp1-26). New York: Routledge.
- Carnoy, M. & McEwan, P. (2000) Does privatization improve education?: The case of Chile's national voucher plan. *Educational Evaluation and Policy Analysis*, 22(3), 213-239.
- Carnoy, M. and Rhoten, D. (2002). What does globalization mean for educational change?: a comparative approach. *Comparative Education Review*, 46(1), 1-9.
- Chubb, J & Moe, T. (1990). *Politics, markets and America's schools*. Washington DC: The Brookings Institution.
- Conklin, J. (1984). *Sociology: an introduction*. New York: MacMillan
- Cook, T. D. (1974). The Potential Limitations of Secondary Evaluations. In Apple, M. W., Subkoviak, M. J. & Lufler, H. S. (Eds.) *Educational Evaluation: Analysis and Responsibility*. (pp155-235). Berkeley, CA: McCutchan.
- Crouch, C. (2011). *The strange non-death of neoliberalism*. Cambridge, United Kingdom: Polity.
- Dale, A., Arber, S. & Procter, M. (1988). *Doing Secondary Analysis*. London: Unwin Hyman.
- Diani, M & Della Porta, D. (1999). *Social movements: an introduction*. Malden, Massachusetts: Blackwell.
- Dohn, N.B. (2007). Knowledge and Skills for PISA – Assessing the Assessment. *Journal of Philosophy of Education*, 41(1), 1-16.
- Figazzolo, L. (2008). *Impact of PISA 2006 on the education debate*. Paper for Education International. Retrieved 12/11/2008 from <http://download.ei-ie.org/docs/IRISDocuments/Research%20Website%20Documents/2009-00036-01-E.pdf>
- Forsey, M. (2007). *Challenging the system? A dramatic tale of neoliberal reform in an Australian high school*. Charlotte, North Carolina: Information Age Publishing.

- Friedman, M. (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.
- Gall, M.D, Gall, J.P, & Borg, W.R. (2010). *Applying Educational Research*. Boston: Allyn and Bacon.
- Goldstein, H. (2004). International Comparisons of student attainment: some issues arising from the PISA stud. *Assessment in Education*, 11(3), 319-330.
- Gorard, S. (2001). *Quantitative Methods in Educational Research*. London: Continuum.
- Gorard, S., See, B. H., Smith, E. & White, P. (2006). *Teacher Supply: The Key Issues*. London: Continuum.
- Gorard, S. & Smith, E. (2004). An International comparison of equity in education systems. *Comparative Education*, 40(1), 15-28.
- Grant, N. (2009). Foreword. In Hill, D. & Kumar, R (Eds.) *Neoliberalism and education and its consequences*. New York: Routledge.
- Green, A. (2006). Education, globalization and the nation state. In Lauder, H., Brown, P., Dillabough, J. and Halsey, A. (Eds.) *Education, globalization and social change*. (pp. 192-197). Oxford: Oxford University Press.
- Grek, S. (2009). Governing by numbers: the PISA effect in Europe. *Journal of Education Policy* 24(1) 23-27.
- Grubb, W. and Lazerson, M. (2006). The globalization of rhetoric and practice: the education gospel and vocationalism. In Lauder, H., Brown, P., Dillabough, J. and Halsey, A. (Eds.) *Education, globalization and social change* (pp. 295-307). Oxford: Oxford University Press.
- Gutstein, D. (2010). War on public schools rages: Right-wing “think tanks” on the attack. *Our Schools/Our selves*, (20)1, 47-54.
- Hakim, C. (1982). *Secondary Analysis in Social Research: a guide to data sources and methods with examples*. London: Allen & Unwin.
- Hamilton, L., Stecher, B., Vernez, G. & Zimmer, R. (2007). Passing or failing: A midterm report card for “No Child Left Behind”. *Rand Review*, Fall 2007, 16-25.

- Hargreaves, A. (2006). Four Ages of Professionalism. In Lauder, H., Brown, P., Dillabough, J. and Halsey, A. (Eds.) *Education, globalization and social change*. (pp 673-691). Oxford: Oxford University Press..
- Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. New York: Teachers College.
- Hanushek, E. and Wößmann, L. (2008). *Education Quality and Economic Growth*. Washington D.C.: World Bank.
- Harvey, D. (2005). *A brief history of neoliberalism*. Oxford: Oxford University Press.
- Heyneman, S. (2003). The history and problems of education policy making at the World Bank 1960-2000. *International Journal of Educational Development*, 23(3), 315-337.
- Hopmann, S.T. & Brinek, G. (2007). Introduction. In Hopmann, S. T., Brinek, G. & Retzl, M. (Eds) *PISA According to PISA: Does PISA Keep What it Promises?* (pp. 9-19). Vienna, Austria: Lit-Verlag.
- Hoult, T. (1969). *Dictionary of Modern Sociology*. New Jersey USA: Littlefield Adams.
- Hursh, D. (2005). Neo-liberalism, markets and accountability: transforming education and undermining democracy in the United States and England. *Policy futures in Education*. 3(1), 3-15.
- Hursh, D. (2007). Assessing No Child Left Behind and the rise of neoliberal education policies. *American Educational Research Journal*, 44(3), 493-518.
- Jacob, B. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics*, 89, 761-796.
- Johnson, B. (2001). Toward a New Classification of Nonexperimental Quantitative Research. *Educational Researcher*, 30(2), 3-13.
- Johnson, B. & Christensen, L. (2000). *Educational Research: Quantitative and Qualitative Approaches*. Boston: Allyn and Bacon.
- Jones, P. (2004). Taking the credit: financing and policy linkages in the economic portfolio at the World Bank, In Steiner-Khamsi, G. (Ed) *The Global Politics of*

- Educational Borrowing and Lending*. (pp 188-200). New York: Teachers College.
- Kamens, D. & McNeely, C. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review* 54(1) 5-25.
- Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of Behavioural Research*. Fort Worth, Texas: Harcourt.
- Kumar, R. & Hill, D. (2009). Introduction: Neoliberal Capitalism and Education. In Hill, D. & Kumar, R. (Eds.) *Neoliberalism and education and its consequences*. (pp.1-11.) New York: Routledge.
- Lather, P. (2004). Scientific research in education: a critical perspective. *British Educational Research Journal* 30(6) 759-772.
- Lauer, P. A. (2006). *An Education Research Primer*. San Francisco: Jossey-Bass.
- Lingard, B. & Grek, S. (2007). *The OECD, indicators and PISA: an exploration of theoretical perspectives*. Paper for the Economic and Social Research Council/European Social Fund retrieved 12/1/12 from [http://www.ces.ed.ac.uk/PDF%20Files/FabQ\\_WP2.pdf](http://www.ces.ed.ac.uk/PDF%20Files/FabQ_WP2.pdf)
- McMillan, J. H. (1992). *Educational Research: Fundamentals for the Consumer*. New York: Harper Collins.
- Marsh, J., Springer, M., McCaffrey, D., Yuan, K., Epstein, S., Koppich, J., Kalra N., DiMartino, M., and Peng, A. (2011). *A Big Apple for Educators: New York City's Experiment with Schoolwide Performance Bonuses*. Santa Monica, California: Rand.
- Martens, K. (2007). How to Become an Influential Actor – The ‘Comparative Turn’ in OECD Education Policy. In Martens, K., Rusconi, A. and Leuze, K. (Eds.) *New Arenas of Education Governance: The Impact of International Organizations and Markets on Educational Policy Making*. (pp 40-56). Basingstoke, United Kingdom: Palgrave MacMillan.

- Meyer, J., Boli, J. Thomas, G. & Ramirez, F. (1997). World society and the nation state. *American Journal of Sociology*. 39(4) 144-181.
- Monkman, K. and Baird, M. (2002). Educational Change in the context of globalization. *Comparative education review*. 46(4), 497-508.
- Moutsios, S. (2009). International organisations and transnational education policy. *Compare*. 39(4) 467-478.
- OECD. (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2007a). *Executive Summary PISA 2006: Science Competencies for Tomorrow's World*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2007b). *PISA 2006: Science Competencies for Tomorrow's World: Volume I - Analysis*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2007c). *PISA 2006 Volume 2: Data*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2009). *PISA 2006 Technical Report*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2010a). *PISA 2009 Results: Overcoming Social background: equity in learning opportunities and outcomes. (Volume II)*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2010b). *PISA 2009 Results: What makes a school successful?: Resources, Policies and Practice (Volume IV)*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2010c). *PISA 2009 Results: What students know and can do: student performance in reading mathematics and science*. Paris: Organisation for Economic Cooperation and Development.
- OECD. (2011). *PISA 2009 Technical Report (preliminary version)*. Retrieved 3/1/2012 from [www.oecd.org](http://www.oecd.org).

- Papadopoulos, G. (1994). *Education 1960-1990: The OECD Perspective*. Paris: Organisation for Economic Cooperation and Development.
- Porter, T. & Webb, M. (2007). *The role of the OECD in the orchestration of global knowledge networks*. Paper for Canadian Political Science Association retrieved 20/12/2011 from <http://www.cpsa-acsp.ca/papers-2007/Porter-Webb.pdf>.
- Prais, S. J. (2003). Cautions on OECD's Recent Educational Survey (PISA). *Oxford Review of Education*, 29(2), 139-163.
- Prais, S. J. (2004). 'Cautions on OECD's Recent Educational Survey (PISA): rejoinder to OECD's response', *Oxford Review of Education*, 30(4), 569-573.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Books.
- Resnik, J. (2006). International organisations, the "education-economic growth" black box and the development of world education culture. *Comparative education review* 50(2), 173-196.
- Rinne, R., Kivirauma, J. & Simola, H. (2002). Shoots of revisionist education policy or just slow readjustment: The Finnish case of educational reconstruction. *Journal of Education Policy*, 17(6), 643-658.
- Rizvi, F. & Lingard, B. (2006). Globalization and the changing nature of the OECD's educational work. In Lauder, H., Brown, P., Dillabough, J. and Halsey, A (Eds.) *Education, globalization and social change* (pp. 247-260). Oxford: Oxford University Press.
- Robertson, S. (2008). Remaking the World: Neoliberalism and the transformation of education and teachers' labour in Compton, M. and Weiner L. (Eds.) *The global assault on teaching, teachers and their unions*. (pp.11-30). Basingstoke, United Kingdom: Palgrave MacMillan.
- Robertson, S., Novelli, M., Dale, R., Tikly, L., Dachi, H. & Alphonse, N. (2007). *Globalisation, education and development: Ideas, actors and dynamics*. London: United Kingdom Department for International Development.

- Sahlberg, P. (2004). Teaching and Globalization. *Managing Global Transitions*, 2(1), 65-83.
- Sahlberg, P. (2007). Education policies for raising student learning: the Finnish approach. *Journal of Education Policy* 22,(2) 147-171.
- Sandström, M & Bergström, F. (2002). School vouchers in practice: Competition will not hurt you! Paper for the Research Institute of Industrial Economics. Retrieved 6/2/2012 from [http://www.ifn.se/eng/publications/wp/2000-2005/2002\\_1/578\\_1](http://www.ifn.se/eng/publications/wp/2000-2005/2002_1/578_1)
- Smith, E. (2007). *Analysing Underachievement in Schools*. London: Continuum.
- Smith, E.(2008). *Using Secondary Data in Educational and Social Research*. Berkshire, England: Open University Press.
- Steiner-Khamsi, G. (2004) Blazing a trail for policy theory and practice in Steiner Khamsi, G. (Ed) *The Global Politics of Educational Borrowing and Lending* (pp.201-220). New York: Teachers College.
- Tilly, C. (2004). Past, present and future globalizations. In Steiner Khamsi, G. (Ed) *The Global Politics of Educational Borrowing and Lending* (pp13-28). New York: Teachers College.
- Turner, R. (2008). *Neo-Liberal ideology: history, concepts and policies*. Edinburgh, Scotland: Edinburgh University Press.
- Thrupp, M. & Hursh, D. (2006). The limits of managerialist school reform: the case of target-setting in England and the USA. In Lauder, H., Brown, P., Dillabough, J. and Halsey, A. (Eds.) *Education, globalization and social change* (pp.642-653). Oxford: Oxford University Press.
- Woodhall, M. (1997). Human Capital Concepts in Halsey, A., Lauder, H., Brown, P., and Wells, A. (Eds.) *Education, Culture, Economy and Society*. (pp.217-223). Oxford: Oxford University Press.
- Wuttke, J. (2007). Uncertainties and Bias in PISA. In Hopmann, S. T., Brinek, G. & Retzl, M. (Eds.) *PISA According to PISA: Does PISA Keep What it Promises?* (pp.241-264). Vienna, Austria: Lit-Verlag.

## APPENDIX

Correlations between level of implementation of GERM policies in participating countries and student performance, PISA 2009, OECD countries

		Science performance	Reading performance	Maths Performance	Predicted reading score for a student with ESCS= 0
Percentage of students in schools where:	Student Results posted publicly	-0.03	0.03	-0.11	-0.02
	School competing	0.08	0.06	0.00	0.16
	Achievement data used for teacher evaluation	-0.30	-0.31	<b>-0.39*</b>	-0.12
	Achievement data used for resource allocation	-0.19	-0.08	-0.32	-0.04
Percentage of students in private schools		0.04	0.05	-0.01	0.09

\*Correlation is significant at the 0.05 level