An Chomhairle Náisiúnta Eacnamaíoch agus Shóisialta
National Economic & Social Council
NESC

Confidential not for Publication

Understanding PISA and what it tells us about educational standards in Ireland

NESC Secretariat Paper[1]

Council Meeting of Friday 16th March 2012

[1] With thanks to Dr. Jude Cosgrove and her colleagues at the Educational Research Centre, and Professor Emer Smyth, ESRI, for points of clarification.

1 Introduction

In first half of the decade, Ireland's average to above average relative performance in the Programme of International Student Assessment (PISA) was generally interpreted as affirming that Ireland's education system was doing well. In 2009, Ireland experienced a sharp decline in its international ranking in PISA scores. This led to concerns being expressed regarding the extent to which investment in education was being translated into improved outcomes and, indeed, 'whether government funds were being put to efficient use' (Newman, 2011:367). Furthermore, with the drop in PISA score rankings suggestive of declines in educational standards, there were also concerns about the wider impact such a decline might have on Ireland's capacity for economic growth and innovation.

This paper explores Ireland's performance in PISA in the past decade and considers the PISA assessment in the context of Ireland's wider system of educational evaluation. Detailed research is undertaken by the Educational Research Centre in Ireland[2] on the factors that contribute to the relative performance of Irish students in PISA. This provides important insights into the possible reasons for the decline in Irish rank performance in PISA 2009. While, this paper draws on this work and, indeed, from that of the OECD on PISA, the focus is less on factors that help explain the relative performance of Irish students in PISA and more on highlighting what PISA can and cannot tell us about student performance and the quality of the Irish education system, as well as the broader implications for policy.

The paper is set out as follows; Section 2 provides an overview of the PISA process. It places PISA in the wider context of Ireland's system of educational evaluation and asks what PISA does and does not tell us. Section 3 provides a short overview of the Ireland's relative performance in PISA. Section 4 considers Ireland's performance and what it says about the quality of Irish education. Section 5 concludes by reflecting on PISA as an indicator of educational performance and on the evaluation of education in Ireland.

[2] The Educational Research Centre, Drumcondra, Dublin, was founded in 1966, and its work comprises implementation and analysis of international and national surveys of education, evaluations of new programmes and education initiatives, test development, and critical analyses of issues in education (www.erc.ie).

2 An Overview of PISA

2.1 PISA: What is it?

Launched in 1997 by the OECD, the Programme for International Student Assessment (PISA) is an international assessment of the knowledge of 15 year olds[3] across three domains - reading literacy, mathematical literacy and scientific literacy. Underpinned by the notion of the knowledge economy, these individual assessments are not intended to capture educational attainment or command of a school curriculum *per se* but rather to act as indicators of the knowledge and skills needed in adult life and for full participation is society. Such knowledge and skills are viewed as central to enabling individuals to take advantage of the globalised world economy (OECD, 2010a:3).

To achieve a quality and equitable educational system, it is both important and necessary for policymakers to understand what factors influence performance. As such, most countries, in some way, monitor and evaluate the quality of their educational systems; student and school performance are one aspect of this. By facilitating the monitoring of certain educational outcomes in the context of an internationally agreed framework, the PISA assessment is considered to provide a useful benchmark for the international comparison of student outcomes. Supporting the analysis of relative performance across countries, it provides a country comparative dimension to the evaluation of educational outcomes which cannot be achieved through national assessment. PISA results offer some insight into the variation in achievement both within and between countries. It also attempts to contextualise the results using information from participants' educational systems, school features, as well as family and individual characteristics (Cosgove & Hislop, 2011). PISA was designed to help governments understand, and thus enhance the effectiveness of, their educational systems. The assessment is widely interpreted as a quality indicator and as tends to be used by governments, policymakers and others as one marker of educational performance(OECD, 2010a:3, De Bartoli & Thomson, 2010:20, 21)

At the same time, the use of PISA, as a basis for policy development, has garnered some criticism; it is argued that differences in culture and language impact on its comparability, that it captures only a subset of skills and that it remains unclear how evidence from PISA can be translated to bring about improvements in national

[3] OECD reporting refers to "15 year olds" as shorthand for the target population. The target population covers students who are aged between 15 years and 3 month and 16 years and 2 months at the time of the assessment and who have completed at least 6 years of formal schooling. See (OECD, 2010a)

education systems (Smyth, E. & McCoy, 2011)[4]. Other criticisms of the assessment include the values underpinning PISA, what PISA measures and is intended to measure, its cultural fairness, as well as its representativeness[5] (Eivers, 2010). On the other hand, it is important to note that such criticisms are not specific to PISA. Many, if not all, equally apply to other international cross-sectional surveys. Indeed, a rigorous set of technical standards are imposed in the development and implementation of PISA, including an improved method of producing and validating translation, as well as vetting and selecting test items on the basis of cultural fairness[6].

2.2 Publishing PISA

The first PISA assessment took place in 2000 and, thereafter, in three yearly cycles, with tests administered in 2003, 2006 and 2009. Each PISA cycle consists of 1 major and 2 minor domain, with the major domain assessed in detail. In 2000, reading literacy was a major domain, followed by mathematical literacy in 2003 and scientific literacy in 2006. The most recently published PISA results are from 2009, with reading literacy once again a major domain. The next PISA assessment will be administered in 2012, with mathematical literacy as the major domain.

In each assessed domain, PISA publishes an average point score for each participating country. The two official measures used to describe a country's performance in PISA is country performance (average point score) relative to the OECD average and relative to each other[7]. In examining country mean score relative to the OECD mean, countries fall into three categories: those whose scores are statistically significantly below, equivalent or above the OECD average. A statistically significant difference suggests that the difference in scores is unlikely to have occurred by chance. To facilitate the comparison of country performance relative to each other, PISA publish, for each country, a list of countries whose scores are equivalent. The published PISA results also include a lower and upper bound of rank performance; it is from this that the widely used headline ranks are derived. The OECD note that it is not possible to determine a precise rank of a country's

[4] See Smyth and McCoy, 2011, pp.3 and 4 for an overview of this literature.

[5] This paper does not address these issues but instead focuses on the direct interpretation of PISA scores and the evaluation of the Irish education system. For an overview of issues to do with the implementation and broader interpretation, see Eivers, 2010.

[6] See PISA Technical Reports. www.oecd.org

[7] The mean and standard deviation in 2000 was 500 and 100 respectively. For PISA 2009, the OECD mean is 493 with a standard deviation of 93 (OECD, 2010a:55). The term the 'OECD average' (or the 'OECD mean' as used in this paper) refers to the mean data values for all OECD countries for which data are available. Each country contributes equally to the average (OECD, 2005:144). It can also be described as the arithmetic mean of respective country estimates where each country is given equal weight in the computation (OECD, 2010a:29)

performance. However, it is possible to determine, with confidence, a range of ranks in which a country performance level lies (OECD, 2010a:55)[8].

The PISA assessment also provides data on the proficiency levels and performance spread of students in each country. Student performance is categorised into 5 main proficiency levels, ranging from the lowest Level 1 to the highest Level 5[9]. Level 5 was expanded in 2009 to create a Level 6[10]. Proficiency levels are determined by grouping student scores on a continuous scale into levels; the cut-points for proficiency levels differ slightly across domains because they are established on the basis of test questions that are specific to each domain. The overall range of student scores for Ireland in 2009 lay between, from lowest to highest, approximately 100 to 800 points in each domain[11]. The OECD provides a description of the skills expected of those who score in each proficiency level. A range of data describing the distribution or spread of student scores includes the standard deviation, as well as percentile scores at the bottom and top percentiles. A successful school system is described by the OECD as combining an above average score with a narrow or below average distribution of student scores. A narrow distribution of scores is interpreted by the OECD as indicating that inequality of learning outcomes is low (OECD, 2010b:13)

In addition to the collection of the domain indicators, a range of background information on students and schools are gathered under PISA[12]. This contextual information is collected to support analysis of student achievement and help to explain and interpret student performance. The information collected about students includes a range of individual characteristics such as gender, level and first language. Context information on home and educational influences is also collected; this includes information such as parental occupational status, parental educational attainment, family structure, family wealth, home educational resources, preschool

[8] this is because the figures are based on samples

[9] A full description of what students in each level are capable of at each proficiency level,, the score cut-point for each level and the percentage of students achieving each level (OECD average and Ireland) is outlined in Perkins et al, 2010 (for Literacy p12, for Maths p19 and for Science p21).

[10] Prior to 2009, PISA proficiency encompassed 5 levels .Level 6 was introduced in 2009 and is in essence an expansion of previous level 5 designed to differentiate the very high performing students (Perkins *et al.*, 2010). 'Level 5 and above' (2009) equates to Level 5 in previous years. Level 1 is also disaggregated to differentiate the very low performers, Level 1a and Level 1b, there is also a 'below level 1b' for those students who did not demonstrate the skill required to answer the easiest PISA reading items. 'Level 1 and below' (2009) equates to Level 1 in previous years.

[11] 93-815 in Reading, 109-780 in Mathematics and 109-855 in Science (Source: ERC)

[12] As part of the PISA process, students complete an extensive background questionnaire, while school principals completed a questionnaire on educational context in their schools.

attendance and school attendance[13]. A further range of indicators pertaining to students attitudes, engagement, motivations and beliefs are collected, as well as those capturing information on the learning environment, that is, on school and classroom climate (De Bartoli & Thomson, 2010: 3)[14]. More specifically, PISA also provides detailed information about students attitude and engagement with reading and computer use, as well as education in general.

2.3 PISA in Practice

PISA is a large cross-sectional study. As such, it captures information about a group of students at one point in time, in essence a snapshot. This supports an examination of the correlation between overall student performance and a range of other variables included in the study[15].

Approximately 470,000 students, from 65 participating countries and representing 26 million 15 year olds , participated in the 2009 PISA assessment (OECD, 2010a:20). In Ireland, recruitment of students to PISA occurred through a random selection of schools, and then pupils (all 15 year olds) from within those schools. Prior to selection, schools were classified by school type, size, gender composition and socio-economic composition to provide a nationally representative sample. Weights were also applied to both the school and student sample to ensure representativeness. Of the 160 schools selected, 144 participated. Up to 35 pupils from each school took part. Some selected students did not participate due to special educational needs, limited experience with language of instruction, non-eligibility due to age rules, as well as, refusals and absences. In total, 3937 Irish students completed the assessment, a response rate of 83.8%[16].

In 2009, the PISA assessment was administered through 13 different test booklets, each book with a subset of PISA questions and one book to each student. This means that each student answers a subset of all questions. There is a systematic overlap in content across booklets. Individual PISA scores are not equivalent to 'percent

[13] Parental occupational status and parental educational attainment are among a number of variables used to determine the socio-economic background of students. The others include home educational resources, cultural possessions in the home and material possessions in the home.

[14] These include a range of questions examining student interest in, engagement and enjoyment in each domain, in addition to information on effort and persistence, beliefs and self-efficacy. Information on aspects of learning and instruction are collected for the major domain of that cycle.

[15] More detailed analysis using longitudinal survey data, such as that carried about the the Economic and Social Research Institute in Ireland, allows more detailed tracking of progress and processes over time. Longitudinal data follows the same students over time and thus allows changes in student performance to be related to changes in other variables.

[16] Weighted response rate

correct' mark but are based on a computation method which, put simply, imputes a student score based on this subset of questions (Perkins *et al.*, 2011:10). The test takes 2 hours. In addition to the test, students answered a 30 minute questionnaire aimed at gathering information about their background, learning habits, attitude to reading, and engagement and motivation. A school questionnaire is also completed by the principal. This is intended to capture information on the context of education at their school, for example, teacher qualifications, number of staff, school and teacher autonomy, resources, policies and practices (De Bartoli & Thomson, 2010: 3, 6)[17].

2.4 PISA in Ireland

A forthcoming NESC Report provides a detailed overview of assessment and evaluation in the Irish education system, briefly considered here is PISA's position within this overall system.

Although not a formal component of Ireland's system of educational evaluation, the 2000, 2003 and 2006 PISA results would seem to have been considered by the policy system as evidence that overall the Irish education system was performing well. It was also heralded by the teaching profession as highlighting the quality of Ireland's teaching force and providing confirmation that that Ireland's schools and teachers delivered a 'high and consistent of standards of education across the school system'[18]. In essence, PISA provided a useful international comparator data source which served to confirm the prevailing view of a high quality Irish education system. Beyond this there was little deeper reflection of PISA's overall contribution to the assessment of Irish educational performance or where it might fit within a broader quality framework for Irish education. Nonetheless, even in these early cycles of PISA, there was some dissatisfaction with performance in mathematics. This was one contributing factor in the development of Project Maths[19], an initiative introduced to develop a revised syllabus in both junior and leaving certificate mathematics[20]. However such domain-specific national responses tend to be influenced by wider contextual or historical factors, for example, with respect to mathematics, there had been not major overhaul of the curriculum since the 1970's.

---

[17] While not a requirement of PISA, a teacher questionnaire was also administered in Ireland in all cycles, this includes information on teaching methods, resources, professional development etc. (www.erc.ie/pisa)

[18] ASTI, International Focus on Irish Education, Volume 26: Number 1, January 2008

[19] www.projectmaths.ie

[20] http://www.ncca.ie/en/Publications/Consultative_Documents/Review_of_Mathematics_in_Post-Primary_Education.pdf - pp. 15-16

Ireland's use of PISA, and indeed the system of education evaluation itself, have, however, been profoundly challenged in a number of respects by the perceived decline in Irish performance in PISA 2009. Firstly, it challenged the simplistic way PISA was used, interpreted and accepted as the headline metric of educational performance. Those involved in the education sector began to ask more searching questions of PISA; does the drop in PISA performance reflect a drop in Irish educational standards? If it does, what factors are influencing student performance? It also prompted questions of the validity of PISA assessment itself. Secondly, and perhaps more importantly, it highlighted the significant information, data and analysis deficit that exists in Irish educational evaluation

The framework for school evaluation in Ireland, developed over the past decade and a half, is a self-evaluation type model. Under the current framework, *Looking at our Schools*, schools are expected to consider their performance on an ongoing basis across five broad areas; quality of school management, planning, curriculum provision, teaching and learning, as well as pupil support[21]. This self-evaluation process also has a related external dimension; it forms the basis of a *Whole School Evaluation,* which includes a detailed school inspection by a visiting Inspectorate every 5 to 7 years[22]. While the evaluation process in Irish school covers a total of 143 'themes for self-evaluation' and results in considerable documentation gathering, culminating in post evaluation verbal and written reports, the analytic capacity of the system as currently configured and implemented is considered extremely limited. For example, the OECD Background Report for Ireland on Improving School Leadership notes that while references are made to quality, no objective evidence is provided in statistical form (OECD, 2007: 13). Although the Department of Education and Science does make use of state examination data, state examinations are not standardised assessments and data is not used for evaluation purposes[23]. In a recent analysis of the Irish self-evaluation system and its implementation, McNamara *et al.* (2011) consider the views of inspectors, education leaders and teachers and highlight a number of weaknesses. They note the process is perceived as one which supports 'impressionistic conclusions' over 'analytic evaluation', seems to be evidence free and lacks hard or usable data[24]. The authors suggest that the analytic capacity of the current system is hampered by a lack of data, the underuse of existing data, as well as a lack of support for, and conduct of, school based research. In short, a systematic evidence based process does not exist.

---

[21] These areas are disaggregated into aspects, components and themes (See NESC, 2012 for a more detailed account)

[22] See McNamara et al,(2011) and NESC (2012) for a more detailed account of this process.

[23] A standardised test is described in Chapter 1 of Shiel *et al* (2010)

[24] McNamara *et al.* (2011) pp70-73

Evaluation and, in particular, analysis of student attainment was historically a particularly contentious issue in Ireland (McNamara *et al.*, 2011, McNamara & O'Hara, 2006)[25]. Although standardised school-based assessments take place at primary level, up to recently, schools were not required to inform the Department of Education and Science (DES) of the outcome[26]. This is to change under the National Literacy and Numeracy Strategy. National assessments using a representative sample are also administered to a sample of primary schools, yet DES cannot identify individual schools. Nevertheless, from an evaluation perspective, there is no systematic national, standardised testing at either primary or post-primary level, where data is collated and examined centrally for system-level evaluation or used by individual school to support internal assessment of a more formative kind. This has led one commentator to conclude that as a result there are 'no accepted benchmarks for the comparison of student achievement and teacher performance' (McNamara *et al.*, 2011: 70). Furthermore, while there are two national examinations carried out in the 3rd and 6th (or final year) of secondary education, the law prohibits the publication of examination results to compare schools or teachers. Ireland is not an exception is this matter. In Denmark, for example, the publication of results, except data aggregated to national level, is prohibited (Shiel *et al.*, 2010). Nonetheless, the debate perhaps should be less about *whether* data should be published and more about *how* it should be published, international examples show ways in which data about schools can be published without ranking them.

A lack of any systematic approach to data gathering and analysis means that extensive data collection does not occur in schools and not enough is done with the data that is collected (such as data on absenteeism, lateness and class assessments). A lack of information on pupil ability at point of entry and information on pupil background also seriously undermines the capacity for meaningful analysis[27].

This is very different to how PISA and, indeed, other national statistical data are used by some other countries in evaluating and assessing the performance of their educational system (Shiel *et al.*, 2010). In Australia, a National Assessment Programme (NAP) incorporates PISA as *one* component of an assessment process encompassing a number of international standardised tests[28] and national literacy and numeracy assessments in years 3, 5, 7 and 9, as well as national sample assessment across a number of specific subject domains in Years 6 and 10. It is

[25] See (McNamara & O'Hara, 2006), for an account of the issues

[26] NESC (forthcoming). Report on Education as part of the Standards and Quality in Human Services Series(NESC, Forthcoming)

[27] For discussion on the data needed, see (Smyth, E. & McCoy, 2009: 226), and (McNamara *et al.*, 2011: 71, 72)

[28] PISA and TIMSS (Trends in International Mathematics and Science Study)

important to note that the NAP itself encompasses only the summative assessment part of Australia's national assessment infrastructure. This is a broad and extensive evaluation framework for education which includes a combination of summative and formative assessment, and is based on both quantitative measuring and qualitative context based appraisal (OECD, 2011a)[29]. In this way, PISA has a formal and structured place in Australia's system of educational assessment and evaluation. The role of PISA is less clear in an Irish context.

2.5 What PISA does and does not tell us

PISA is a complex and ambitious assessment with a significant focus on the application of skills in literacy, numeracy and scientific understanding. While the test does provide important and useful information regarding the accumulation of knowledge and skills, the publication of its results in the form of league tables means that it can be reported in an overly simplified manner (Cosgove & Hislop, 2011) (Lowell & Salzman, 2007: 16). It is important that in focusing on PISA results, some attention is given to the way the data is published and how it should be interpreted.

Firstly, PISA cannot identify cause-and-effect relationships between inputs, processes and educational outcomes but can highlight key features in which education systems are similar and different. PISA results demonstrate what is possible and what can be achieved in education as indicated by the highest performing countries. In this way, it can be used by policymakers to benchmark the knowledge and skills of students in their own country with those in other countries. Furthermore, it supports an examination of performance change compared with that observed elsewhere, telling us something about the pace of educational progress (OECD, 2010a: 20)[30].

Secondly, a difference in scores between countries does not automatically imply that schools or aspects of the educational system in one country are more effective than then other, what it does imply is that "cumulative impact of learning experiences, starting in early childhood up to the age of 15 and embracing experiences both in school and at home, have resulted in higher outcomes in the literacy domains that PISA measures"(OECD, 2004a: 320) (Lowell & Salzman, 2007: 18). If the PISA assessment provides an indicator of the whole-life learning development of 15 year olds and reflects a range of factors combining country, school, classroom, family and

[29] For a detailed overview and assessment of the Australian evaluation system see the OECD Review of Evaluation and Assessment in Education, 2011

[30] PISA can and is being used to examine the similarities and differences in country performance at both a point in time and over time, to set policy targets against measureable goals achieved by other system, to initiate research and peer learning designed to identify policy levers and to reform trajectories for improving education

individual effects, then efforts to improve both quality and equity through public policy requires a *long-term* view and a *broad* perspective(Lowell & Salzman, 2007: 16). For example, policies implemented now, such as provisions for pre-school education, will not show dividends, as measured by PISA, until the relevant cohort (those at the receiving end of such a policy) reach 15 years of age. Similarly, a multiple of factors may contribute to poor performance, so while policies aimed at the educational system may support improved performance, policies related to other areas such as social protection may also impact.

Thirdly, countries may differ not just in mean performance but also in the distribution of achievement. PISA data facilitates the comparison of performance in reading, science and maths literacy of Irish students with those of other countries; it also supports an examination of trends in the performance of Irish students over time. While the primary focus tends to be on mean performances, countries also differ in extent to which the variation in their mean score might be due to distribution of PISA scores across the student population. For example, an overall high score can be produced not only because of exceptional performance in the top percentiles but because the disparity between the lower and middle performers is relatively small. Thus the examination of the distribution of achievement can be of significant interest from a policy perspective. PISA scores are set to an OECD average of 500 with a standard deviation of 100[31]; the standard deviation refers to distribution or spread of the scores. Thus while two countries might have similar average achievement, a different distribution of scores (that is a larger standard deviation) may indicate less equitable outcomes. Even where two countries return similar means and standard deviations, there may remain differences in country scores at the lower and/or upper ends of the distribution. As such, performance at 10^{th} or 90^{th} percentile indicate the differences in scores among the very low achievers and/or the very high achievers (Perkins *et al.*, 2010: 6).

Fourthly, a simplistic reading of country rank can lead to misinterpretation and misuse of the data. Official reporting of PISA results are communicated as cross tables of the mean value by country, indicating whether the mean score difference are statistically significant (that is, the difference in scores is unlikely to have occurred by chance). Focus tends be on the mean value by country, reported in the form of country league tables ranked by performance. A ranking is not an absolute score but a relative score. Thus, focusing on country rank, one common (mis)interpretation of PISA scores is that a drop in ranking equates to a decline in performance**.** PISA rankings are influenced by the number of countries that participate in the PISA assessment, this number has changed over time. In 2000, 32

[31] A standard deviation of 100 means that, on average across the OECD, 66% of students score between 400 and 600, and 95% between 300 and 700.

countries participated in the PISA assessment, with 7 more taking the test in 2001. In 2009, this has risen to 65 (with 10 more taking it in 2010). A drop in rank will automatically occur if countries that newly participate in PISA score higher. Thus, a drop does not necessarily (but still might) reflect a drop in performance. This problem of interpretation can be overcome if it is assumed there has been no change in the number of countries participating over time and compare performance of only those countries that participated at two separate time points, for example 2000 and 2009. In this way, the impact of additional participants on the ranking are removed. Nonetheless, ranks still represents a relative as opposed to an absolute score. Thus, a country might do the same or better academically in 2009 compared with 2000 but drop in rank because other countries may have done better. Considering a country's performance over time, the OECD compares the performance of each country to the OECD average. This is not a comparison of Ireland's performance in 2003 and Ireland's performance in 2009, but a comparison of Ireland's performance to that of the OECD average in 2003 and Ireland's performance to that of the OECD average in 2009. This means that we do not actually obtain a comparison of a countries performance to itself over time in PISA trend results. For example, in describing Ireland performance in maths between 2003 and 2009, a focus on rank would highlight Ireland's drop in rank from 20th to 26th. While a comparison of the performance of Ireland to the OECD average across time would describe Ireland as having an equivalent score in maths to the OECD average in 2003 compared to 2009 when Ireland scored significantly below the OECD average[32].

One suggested counterpoint to the ranking issue, however, is that *ranks matter, relativity matters*. If we are competing with other countries for foreign investment and foreign investors care about the relative skills of a country's workforce, Ireland's performance relative to that of other countries, and not Ireland's absolute score, might be the only information of interest[33]. From this perspective, a decline in Ireland's relative score, regardless of whether that represents an absolute decline or not, might be something to worry about. At the same time, whether, or the extent to which, country performance in PISA is considered in location decisions of employers or multinationals is debatable. And if it is, what is the role of other metrics such as the percentage of young people with third-level qualifications?

Fifthly, it is not clear how to interpret the gap in scores between countries or their importance. The interpretation of the meaning of differences in PISA scores is the

[32] Bearing in mind the OECD average in 2009 is not the same as in previous cycles. This is due to new countries joining or where country results are not deemed sufficiently comparable and are excluded. See (OECD, 2010c:136) for a detailed account of this.

[33] This point is made by Kevin Denny, School of Economics, U.C.D: Ireland's PISA results; myths and reality. (http://kevindenny.wordpress.com)

subject to discussion among PISA researchers and others. When citing the rankings, this analysis tends to receive little attention among policy makers and others. The raw PISA data, the absolute test scores, undergo a conversion. This conversion is a technical and complex process based on weighted measures of different raw score components[34]. The data is normalised so that the mean score is 500 and a standard deviation is 100.

Some commentators have argued that while this creates a population distribution of scores, it does not tell anything about the extent of the actual differences in the test results (Lowell & Salzman, 2007: 20). They conclude, from this information, it is difficult to know what a difference in PISA scores represents. That is, does it represent a small or large difference in actual scores? Elaborating on this point in respect of the US, Lowell and Salzman (2007) were prompted to ask "Does the level of panic about lagging US performance, and characterisations of a student population falling dramatically behind those of other countries correspond to actual performance differences of a few percentage points?". On the other hand, if a score distribution has a known standard deviation, score difference can be interpreted with respect to standard deviation units and percentiles. In short, typical score difference can be generated. Proficiency levels and years of education are two ways in which differences in PISA scores have been represented by both the OECD and other bodies who specialise in the analysis of PISA data. The Education Research Centre suggests that 80 points equate approximately to 1 proficiency level difference. Another way in which score differences have been represented relates to years of education, the OECD estimate that a 40 point difference (half a proficiency level) equate to approximately 1 year of education (OECD, 2010a, Perkins *et al.*, 2010).

Sixthly, the implementation of a standardised test across many countries is a complex task. As such, contextual and cultural differences may be interpreted to account for some of the difference in tests scores. The literature points to the difficulties in cross country comparison given the inherent differences in language, culture and national curriculum (Smyth, E. & McCoy, 2011)**.** Eivers (2010: 102) questions the assessments cultural fairness on a number of grounds; the quality and equivalence of the test translations, the potential Anglophone bias, the difference by country in how students respond to items and the differences in the importance given to the assessment by students.

[34] Indeed, one commentator notes that an 'unfortunate by-product of the complexity of the statistical techniques used in PISA is that few feel qualified to debate what PISA does and what it means' (Eivers, 2010)

2.6 PISA as a Systemic Indicator

Ireland is not the only country to be dissatisfied with their performance in PISA and seeking to improve. Thus it is important from a policy perspective that the outcome of the assessment is properly understood and there is a balanced examination of what the data is telling us.

The 2006 NESC Strategy Report elaborates on differences between systemic, diagnostic and performance indicators. PISA is probably best described as a systematic, high-level indicator which gives an overall picture of how a system is performing. It might also be argued that the more detailed contextual components of PISA could, in specific circumstances and as a component of analysis using a multiplicity of evidence, be employed to support a diagnostic analysis (NESC, 2005: 160). National assessments which are sample based can be used for diagnostic purposes at the system level, while census-based national assessment can be used for both diagnostic and performance monitoring (Shiel *et al.*, 2010:34, 35).

For all assessments, the importance of using and interpreting such data in the correct and appropriate way should not be under-estimated. In a recent review of standardised testing, Shiel *et al* (2010:36) note that it is important for users to be aware of limitations of tests, as well as the undesirable, if unintended, consequences of their use. In the development and implementation of national assessment programmes, the possibilities for misuse, intentionally or otherwise, need to be safeguarded against[35].

3 Exploring and Interpreting Ireland's Performance in PISA[36]

Up to 2009, Ireland's performance in PISA across both subject domains and years could be categorised as 'average to better than average'. By 2009, it appeared that the pendulum had begun to swing in the opposite direction. By which ever measure – ranks, scores, and score relative to the OECD average – Ireland's overall performance in PISA seemed to have declined. However the magnitude of the decline, what it meant, and how it should be interpreted, was less clear.

Each PISA cycle consists of 1 major domain and 2 minor, in reporting performance over time, the OECD compares (to ensure reliability) each domain to when it was last

[35] See Shiel *et al* (2010), Chapter 4, for a discussion of the consequences of low and high stakes testing.

[36] Parts of this section draw heavily from an array of primary data analysis carried out by the OECD and the Education Research Centre in Ireland

a major domain, Reading 2000 versus 2009, Maths 2003 versus 2009 and Science 2006 versus 2009.

3.2 Reading Literacy

A Snapshot of Performance in 2009

In the 2009 PISA assessment in reading literacy, Ireland's headline score of 496 represented an overall rank of 21st of 65 participating countries, 17th of 34 OECD countries and 17th of those 39 countries who had participated in both years. Overall, of the other participating economies or countries, the five highest performing were Shanghai –China (556) , Korea (539) , Finland (536) , Hong Kong- China (533) and Singapore (596). With Korea (539) , Finland (536), Canada (524), New Zealand (521) and Japan (520), the top five among the OECD countries .

In examining country mean score relative to the OECD mean, countries fall into three categories: those whose scores are statistically significantly below, equivalent or above the OECD average. Ireland's score of 496 puts it in the middle group, those countries whose performance did not differ significantly from the OECD average, along with 26% of the participating countries. This puts Ireland's performance on a par, relative to the OECD average, with that of the US, France, Germany, Sweden, UK, Denmark, Portugal and Hungary[37]. Based on this measure, 38% of the participating countries scored above Ireland; these included countries such as Finland, Canada, Korea, Netherlands and Norway. While 35% of OECD countries (encompassing Italy, Spain, Greece, Austria, Luxembourg) scored below.

Literacy 2009: Score of OECD Countries Relative to the OECD Average

OECD Average	%	OECD Countries (34)
Above	38%	Korea, Finland, Canada, New Zealand, Japan, Australia, Netherlands, Belgium, Iceland, Norway, Estonia, Switzerland, Poland
At the average	**26%**	US, Sweden, Germany, **Ireland**, France, Denmark, UK, Hungary, Portugal,
Below	35%	Italy, Slovenia, Greece Spain, Czech Rep, Slovak Rep, Israel, Luxembourg, Austria, Turkey, Chile, Mexico

Consideration of Ireland's mean score relative to that of all other participating countries shows that Ireland's mean score does not differ significantly from the

[37] Digital literacy (a computer based assessment of reading) was also examined in 2009. 19 Countries participated including Ireland. Ireland's mean score was 509, a score that was significantly above the OECD average (ranking 8th of 19th)(Cosgrove, J. *et al.*, 2011).

scores of 16 other countries, the US, France, Germany, Sweden, UK, Denmark, Portugal and Hungary, in addition to Norway, Estonia, Switzerland, Poland, Iceland, Liechtenstein, Chinese Taipei (OECD, 2010a: 54)[38].

In the years where literacy was a major domain, 2000 and 2009, information on detailed performance subscales are available; these are Access and Retrieve, Integrate and Interpret, Reflect and Evaluate, Continuous and Non-Continuous scales. Overall, the data shows little variation in average Irish performance across these subscales in 2009. In one category, Reflect and Evaluate, Ireland's mean score was significantly above the OECD average, across the remaining subscales Ireland performance did not differ significantly from the OECD average.

Proficiency Levels and Performance Spread in 2009

In Reading Literacy, just over 17% of Irish students fall into the lowest performing category, Level 1 (or below). This is closely comparable to the OECD average of nearly 19% and to other countries with similar overall performance means such as the UK at 18.4%[39]. Nonetheless, there are countries who do considerably better with only 8.1% of students in Finland and 10.3% of students in Canada falling in this low achieving bracket. An examination of the higher levels shows that Ireland again has a comparable level of highly skilled readers (7.1% of those achieving Level 5 or over) to the OECD average (7.5%) , UK (8%) and Germany (7.6%) but has a much lower level than Finland (14.5%) and the US (9.9%) (Perkins *et al.*, 2010, OECD, 2010a)[40]. Many countries, with mean scores considerably higher than Ireland's, have lower percentages of students in the low performing categories and a higher percentage falling in the high proficiency categories (for example, Korea, Finland and Canada). New Zealand, one of the top 5 OECD countries, mirrors Ireland's performance in the low to mid proficiency levels but returns considerably higher percentages of students in the high performing levels[41]. Finland, on the other hand, combines a considerably

[38] Note that country score relative to the OECD average and score relative to comparison countries are two separate measures. For example, while Italy's score of 486 and Mexico's score of 425 place both these OECD countries in the 'below the OECD average' category, there is a statistically significant difference in their scores when they are compared to each other. A number of countries score statistically significantly above the OECD average yet also return a score that does not differ statistically from the Irish score (Norway, Estonia, Switzerland, Poland).

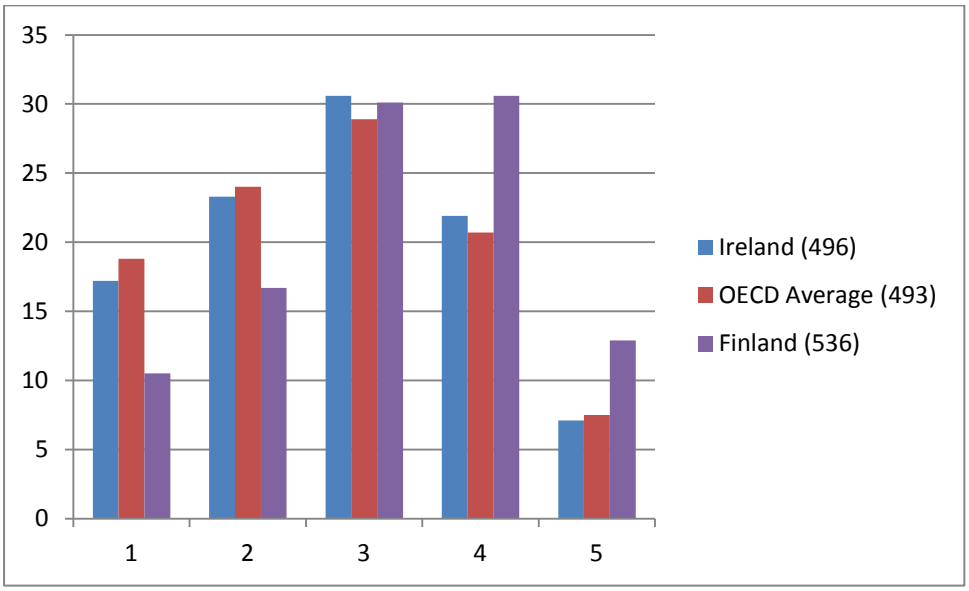
[39] Those at level 6 [score] higher than 698 points, level 5 encompasses scores of higher than 626 but lower than or equal to 698 points, at Level 4 higher than 553 but lower than or equal to 626, at level 3 higher than 480 but lower or equal to 553, at level 2 higher than 407 but lower than or equal to 480, Level 1a higher than 335 but lower than or equal to 407, level 1b higher than 334 but lower than or equal to 262, and there is also a below level 1 b where PISA does not assess the skills of students.

[40] Data source OECD, 2010a, p194

[41] See (OECD, 2010a:194)

lower number of students in the lower proficiency levels with considerably higher numbers in the higher proficiency levels.

Reading: Percentage of Students in each Proficiency Level: Ireland, OECD Average, Finland



***Data Source: OECD, 2010a, p194**

While some countries combine a high average score with a narrow gap between the low and high performers (Korea and Finland), this is not true of all countries. There is wide variation in the distribution of performance for each country, with the OECD noting that the gap between the high and low performers does not seem to be associated with overall level of performance. Both high (Korea) and low (Chile) performing countries show a narrow distribution of performance[42], while some with a wide distribution of performance score well above the OECD average (New Zealand) and others score well below (Qatar) (OECD, 2010d: 53). Finland (536) and New Zealand (521) are both high average performers with significantly above average scores, yet there is a difference of over 40 points between them in the gap between the 10th and 90th percentile. So while Finland and New Zealand return similar high average performances, New Zealand shows a wider spread of performance around the mean and thus is less successful than Finland in achieving equality of outcomes. While the gap between the 10th and 90th percentile in Ireland (238) is somewhat higher than Finland's (223), it is much lower than New Zealand's (266). Finland scores over 30 points higher than Ireland at the 10th, 50th and 90th percentile, with even higher differences at the lower end of distribution[43]. This results

[42] As indicated by the gap between the 10th and 90th percentile (OECD, 2010a: 197)

[43] Indeed, Sabel et al, in their examination of Finnish schools, note that, overall, in PISA the bottom 20% outscore the bottom 20% elsewhere (Sabel *et al.*, 2011)

in both a high mean score and a narrow distribution of performance. The achievement of high scores at the upper end of distribution pushes New Zealand's mean score well above that of Ireland, but also results in a larger gap between the 10th and 90th percentile. The distribution of performance (and the gap between the low and high achievers) in Ireland mirrors that of the OECD average.

In essence, high average performance is only one part of the PISA story, the distribution is also important both in terms of the gap between the low and high achievers, but also in terms of the relative performance of low and high achievers with their counterparts in other countries.

Gender Differences in Reading Literacy

In every country, including Ireland, females outperformed males in the assessment of reading proficiency. Irish females achieved a mean score of 515 compared to a score of 476 among Irish males, a difference of 39 points. This compares with the OECD average score of 513 for females and 474 for males, also a difference of 39 points. Putting this in context, of all participating countries, Columbia had the smallest gender difference (9 points) and Albania the highest (62 points). Interestingly, Finland, a high scoring country with a relatively narrow distribution of performance, returns a 55 point gender difference, the highest of the OECD countries. Chile has the lowest at 22 points. (OECD, 2010a: 197) (Perkins *et al.*, 2010: 12)

Reading Literacy: Changes between 2000 and 2009

Reading was a major domain both in PISA 2000 and 2009. Of the 39 countries that participated in both years, Ireland's score of 496 in 2009 amounts to a drop of 31 points, down from 527 in 2000[44]. This was the largest drop recorded across all of the 39 countries who participated in both 2000 and 2009 (the next largest drop was 22) and represents a drop in rank from 5th to 17th among those countries. An examination of Ireland's performance relative to the OECD average in 2000 and 2009 shows that Ireland score was statistically significant above the OECD average in 2000 but did not differ significantly from the OECD average in 2009. (Perkins *et al.*, 2010:13). The decline in Ireland's performance was uniform across ability levels and thus reflects a drop in relative performance at both the high and lower end. The percentage of students performing in Level 1 or below (the lowest proficiency levels) increased by approximately 6 percentage points, rising from 11% to 17.2%[45].

[44] The 31 point drop includes a decline of 11 points between 2000 and 2003.

[45] It is important to note that while proficiency levels differ slightly across domains, they remain the same over time within each domain, although in 2009 both level 1 and level 5 were both further disaggregated to capture both the lowest and highest performing.

Similarly, in the highest level 5 or above, the percentage of students decreased by approximately 7 percentage points, from 14.2% to 7%. (Perkins *et al.*, 2010:14).

While Ireland had a much lower percentage of students performing at Level 1 and a much higher percentage at Level 5 than the OECD average in 2000, the percentage of Irish students in both the low and high proficiency levels mirrored that of the OECD average in 2009[46]. An examination of Ireland's performance at key percentiles shows that the 30 point drop in Ireland's mean score between 2000 and 2009 is mirrored at 10th, 25th, 75th and 90th percentile points.

There is a marked increase in the percentage of males achieving a proficiency level below level 2, up 10% compared with 3% for females. There has also been an overall increase in the gap between males and females from 29 to 39, this increase in favour of females is also reflected across other OECD countries (Perkins *et al.*, 2010).

3.3 Mathematics Literacy

A Snapshot of Performance in 2009

In the 2009 PISA assessment in mathematics literacy, Ireland's score equated to a rank of 32nd of 65 participating countries and 26th of 34 OECD countries. Overall, of the participating countries the five highest performing were Shanghai –China (600), Singapore (562), Hong Kong- China (555), Korea (546) and Chinese Taipei (543), with Korea (546), Finland (541), Switzerland (534) Japan (529), and Canada (527), the top five among the OECD countries.

The mean score for Ireland in mathematics was 487; this score was below the OECD average of 496. Focusing on the mean value by country relative to the OECD mean, Ireland's score puts it in those group of countries whose scores are below and statistically significantly different to the OECD average. Other countries in this group include the US, Portugal and Spain. Overall 67% of countries had scores that were above or equivalent to the OECD average. Countries with mean scores equivalent to the OECD average include Austria, France and Poland (Perkins *et al.*, 2010: vii)

[46] See Table 2.5 Perkins

Mathematics 2009: Score of OECD Countries Relative to the OECD Average

OECD Average	%	OECD Countries (34)
Above	41%	Korea, Finland, Switzerland, Japan, Canada, Netherlands, New Zealand, Belgium, Australia, Germany, Estonia, Iceland, Denmark, Slovenia
At the average	26%	Norway, France, Slovak Rep, Austria, Poland, Sweden, Czech Rep, UK, Hungary
Below	**32%**	Luxembourg, US, **Ireland**, Portugal, Spain, Italy, Greece, Israel, Turkey, Chile, Mexico

Consideration of Ireland's mean score relative to that of other countries shows that Ireland's mean score does not differ significantly from the scores of 10 other countries, Sweden, UK, Czech Republic, Hungary, Luxembourg, the United States, Portugal, Spain, Italy and Latvia (OECD, 2010a: 134).

Both mathematics and science were minor domains in 2009, this means that the data is limited to a single overall scale score for each domain and cannot be categorised into subscales with any reliability.

Proficiency Levels and Performance Spread in 2009

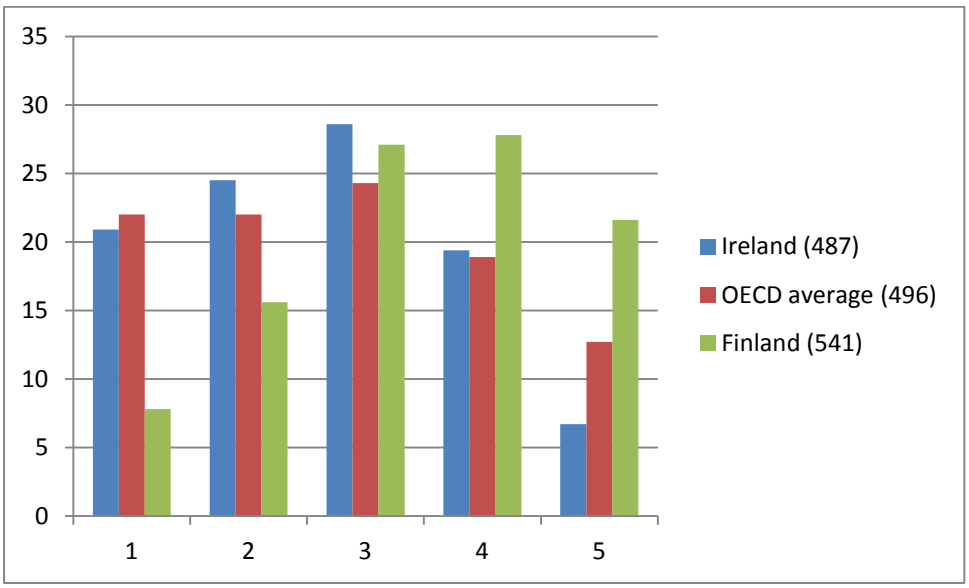
In Mathematics Literacy, just under 21% of Irish students fell in the Level 1 or below[47], indicating a very low performance in maths for one fifth of students. While this is similar to the OECD average of 22%, it represents a slightly better score than other countries with similar overall performance means (and also below the OECD average) such as the US, Portugal and Spain at just over 23%. However, Ireland has a notably lower percentage of student falling in the high performance levels (level 5 or over) to the OECD average with only 6.7%, compared to an OECD average of 12.7%. Ireland also scores lower than other countries with similar overall performance means, the US (9.9%), Portugal (9.6%) and Spain (8%).

Those countries with mean scores that are significantly higher have considerably lower percentages of student falling in the low proficiency categories, the top OECD performers Korea and Finland have 8.1% and 7.8% of student respectively falling in these low achieving categories. They also have a considerably higher percentage of students achieving a high performance score of level 5 or above, Korea 25.5% and Finland 21.6%. 3.1% of students on average across the OECD performed at level 6, with around 8% of students from countries such as Korea and Switzerland achieving

[47] The lower cut-off point for each level of mathematics proficiency in 2009 was 669 for level 6, 607 for level 5, 545 for level 4, 482 for level 3, 420 for level 2, 358 for level 1

Level 6 and more than 5% in countries such as Belgium, Japan and New Zealand. Among some of the non-OECD countries, over 10% of students performed at level 6 (Singapore, Chinese Taipei and Hong Kong-China) while in Shanghai-China over a quarter of students achieved at level 6 performance. In contrast, Ireland with countries such as Mexico, Chile and Greece less than 1% of student managed a Level 6 score. This suggests that Ireland's low average performance is partly due to the comparatively low performance in the high-achieving categories (Perkins *et al.*, 2010, OECD, 2010a).

Mathematics: Percentage of students in each Proficiency Level: Ireland, OECD Average, Finland



***Data Source: OECD 2010a, p221**

As with Literacy, Finland has one of the highest mean performances in mathematics, of the OECD countries, in addition to one of the narrowest distributions[48]. The gap between the 10th and 90th percentile is somewhat smaller than the other top performing OECD countries such as Switzerland, Japan and Canada. While these countries match Finland in the 90th percentile score, they achieve score of between 15 and 30 points lower at the 10th percentile mark. This impacts negatively on their overall mean performance. Among the high performing partner countries and economies, a high mean score is coupled with a large gap between the 10th and 90th percentile. In the main, this is due to some very high scores at the 90th percentile. While, the gap between the 10th and 90th percentile in Ireland is low, equating to that of Finland's, overall Ireland scored approximately 50 points lower than Finland at each key percentile (10th, 25th, 50th, 75th 90th). Interestingly, while Ireland mirrors the OECD average score at the 10th and 25th percentile, it returns lower scores at the 75th

[48] Between 10th and 90th percentile

and 90th,. Ireland scores a little higher at the 10th percentile and a little lower at the 90th percentile than the other below average performers with similar means(OECD, 2010a: 133, 224). It is noted that Ireland's low average performance is partly due to the low relative proportions of students in the high proficiency levels (Perkins *et al.*, 2010:19). Nonetheless, taking the distribution of performance into account, an equal point score improvement at each key percentile would serve to increase Ireland's average score and, at the same time, maintain the narrow gap between high and low performers.

Gender Differences in Mathematics Literacy

In the mathematical literacy domain in 2009, males were higher performers than females in Ireland on mean score, 490.9 versus 483.3. While not statistically significantly different from each other, both scores were significantly below the OECD average. Of the 34 OECD countries, 21 had a significant gender gap with males outperforming females. Belgium experienced the largest difference of nearly 22 points. In Ireland, about 20% of both males and female experienced low levels of performance (Level 1 or below), similar to the OECD average for both genders. However, more proportionately males than females score in the high achieving levels (Level 5 and above), as they do on average across the OECD. Nonetheless, Ireland compares poorly to the OECD average with 8.1% of males and 5.1% of females achieving at high level versus 14.8% and 10.6% respectively (Perkins *et al.*, 2010).

Mathematics Literacy: Changes between 2003 and 2009

Comparing 2003 and 2009, Ireland's rank in mathematics dropped from 20th to 26th among countries that had participated in both years[49]. The decline in Ireland's score was 16 points, the second largest decrease among these countries. Perkins et al (2010) note that while the Irish performance declined slightly between 2003 and 2006, the majority of the decline occurred from 2006 onwards (14 of the 16 points). Ireland's performance was equivalent to the OECD average in 2003 but dropped below in 2009. In 2003, Ireland had significantly fewer students in the low achieving groups (16.8%) compared to the OECD average (21.5%). By 2009 this had changed with no significant difference to the OECD average recorded (Ireland 20. 8%, OECD, 22%). Conversely, the percentage of students achieving at Level 5 or above decreased from 11.4% to 6.7%. This is almost 50% below the OECD average of 12.7% in 2009 and represents a statistically significant decrease. Both the performance of males and female dropped significantly between 2003 and 2009. Males did better then female in both years but only significantly so in 2003. Overall, the decline in Maths performance is considered evenly distributed, although it was greater for high

[49] 39 countries. (OECD, 2010c)

achievers, 6.7% at level 5 or above in 2009 compared with 11.4% in 2003 (Perkins, 2010:20).

3.4 Science Literacy

A Snapshot of Performance in 2009

In the 2009 PISA assessment in scientific literacy, Ireland's mean score of 508 equated to a rank of 20th of 65 participating countries and 14th of 34 OECD countries. Overall, of the participating countries the five highest performing were Shanghai – China (575), Finland (554), Hong Kong- China (549), Singapore (542) and Japan (539), with Finland (554) , Japan (539), Korea (538), New Zealand (532) and Canada (529) the top five among the OECD countries.

Ireland's mean score of 508 lies above the OECD average of 501. Focusing on the mean value by country relative to the OECD mean, Ireland's score puts it in the group of countries whose scores are above and statistically significantly different to the OECD average. There are 15 other countries in this group, in addition to the top OECD performers cited above, these include Estonia, Australia, the Netherlands, Germany, Switzerland, the United Kingdom, Slovenia, Poland and Belgium. A further six OECD countries had mean score equivalent to the OECD average, Hungary, the US, the Czech Republic, Norway, Denmark and France(OECD, 2010a: 150).

Science 2009: Score of OECD Countries Relative to the OECD Average

OECD Average	%	OECD Countries (34)
Above	**44%**	Finland, Japan, Korea, New Zealand, Canada, Estonia, Australia, Netherlands, Germany, Switzerland, UK, Slovenia, Poland, **Ireland**, Belgium
At the average	18%	Hungary, US, Czech Republic, Norway, Denmark, France
Below	38%	Iceland, Sweden, Austria, Portugal, Slovak Rep, Italy, Spain, Luxembourg, Greece, Israel, Turkey, Chile, Mexico

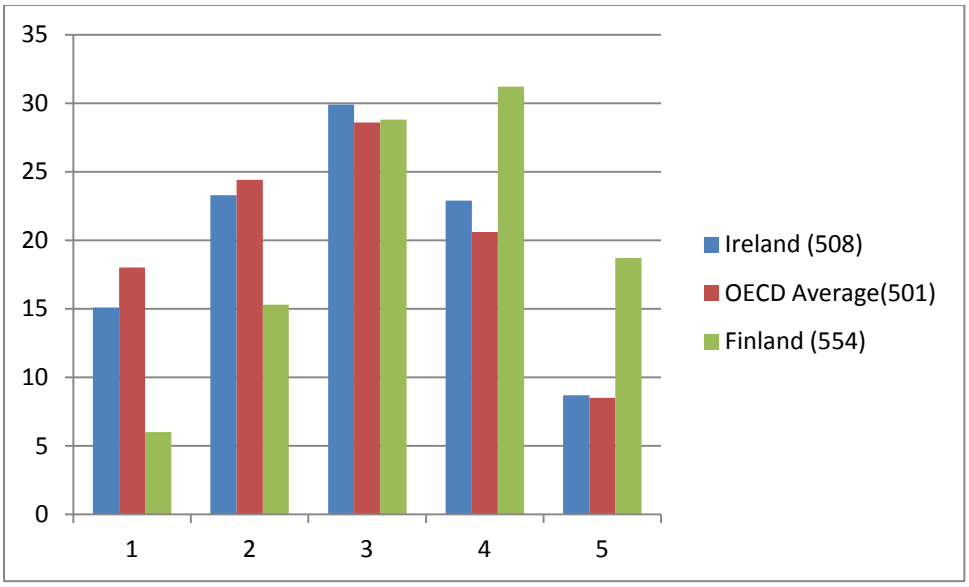
Consideration of Ireland's mean score relative to that of other participating countries shows that Ireland's mean score does not differ significantly from the scores of 9 other countries, UK, Czech Republic, Hungary, the United States, Slovenia, Macao-China, Poland and Belgium (OECD, 2010a: 151).

Both mathematics and science were minor domains in 2009, this means that the data is limited to a single overall scale score for each domain and cannot be categorised into subscales with any reliability.

Proficiency Levels and Performance Spread in 2009

In Science, the percentage of student at Level 1 or below is significantly lower than the OECD average, 15.2% to 18% respectively. Nonetheless, other countries who also achieved mean scores significantly higher than the OECD average show an even lower percentage of students in the low proficiency levels than Ireland, for example Canada at 9.6% and Estonia at 8.3%. The percentage of Irish students at high achieving levels (8.7% who are 5 or over) is not significantly different to the OECD average of 8.5%. Nonetheless, the top performer, Shanghai-China has 24.3% of students in the high achieving category, while Finland, with the top OECD performance had 18.7% of students at this level.

Science: Percentage of Students in each Proficiency Level: Ireland, OECD Average, Finland



***Data Source, OECD, 2010a, p225**

Narrow distributions in performance are evident both among the highest and lowest performing countries[50]. Shanghai-China and Korea combine high mean performances with narrow distributions (a gap of 207 and 209 score points respectively between the 10th and 90th percentile). While Korea's mean performance is almost identical to that of Japan and Singapore, Korea has a much narrower distribution (50 to 60 point less), scoring approximately 25 points higher at the 10th percentile and 20 points

[50] Among the OECD countries the narrowest distributions between the 5th and 95th percentiles were in the lowest performing countries such as Mexico, Turkey and Chile (OECD, 2010a: 150)

lower at the 90th percentile. Among the top performers New Zealand returns the highest gap between the 10th and 90th percentile, its high score at the 90th percentile explains its high overall mean relative to Ireland's. The disparity in the gap in performance across the top performing countries in science, both OECD and non-OECD, is striking. Some countries do very well at combining high scores with narrow gaps, while other achieve high mean scores by balancing lower scores at the 10th percentile with higher scores at the 90th. Although Ireland falls in the 'above OECD average' category, its performance at the 10th, 50th and 90th percentile is very similar to the OECD average performance[51].

Gender Differences in Science Literacy

While females achieved a slightly higher mean score point in Science Literacy, the difference was not statistically significant (females 509.4 and males 506.6). Both scores are above the OECD average of 500.9 for females and 500.8 for males but while the females score for Ireland differs significantly from the OECD average, the score for males does not. Overall, gender differences are small. Among OECD countries, the US (14 points) and Denmark (12 points) have the largest difference in favour of males, while Finland (15 points) and Slovenia (14 points) in favour of females. In Ireland there are slightly more males at the lower proficiency levels (16% versus 14.3%), both percentages below the OECD average. Ireland also has more males in the high achieving categories (9% versus 8.3%), with both almost identical to the OECD average (Perkins *et al.*, 2010: 22).

Science Literacy: Changes between 2006 and 2009

Between 2006 and 2009, Ireland's mean score for science did not change, 508.3 and 508 respectively, and remained significantly above the OECD average. Ireland's rank remained static among OECD countries and stood at 14th of 34. It rose from 18th to 20th among those 57 countries participating in both cycles. The percentage of students in both high and low achieving categories did not change in Ireland since 2006, standing at 15.1% and 8.7% respectively in 2009. The gender gap remained small and insignificant.

[51] (OECD, 2010a: 228)

3.5 Ireland's Performance in Context

Interpreting Ireland's Performance in PISA

An examination of Ireland's performance in PISA by ranking, scores and score relative to the OECD average shows that Ireland is doing less well in the PISA assessment than in previous years and, indeed, that its performance in PISA relative to other countries has declined.

	PISA Rankings for Ireland*			**Actual PISA Scores for Ireland**			**Relative to the OECD average****		
	Literacy	Maths	Science	Literacy	Maths	Science	Literacy	Maths	Science
2000	5th			527			Above		
2003		20th			503			At	
2006			18th			508			Above
2009	17th	26th	20th	496	487	508	At	Below	Above

*among countries participating in both years

** 34 OECD countries

However, it is important to put the results in context. The reporting of raw country ranks is a simplistic and potentially misleading account of PISA results. The magnitudes of score differences, as well as measurement error, need to be taken into account. For example, although Ireland ranked 17th out of 34 OECD countries in reading literacy in 2009, just 8 of the 16 higher-scoring countries had mean scores that were statistically significantly higher. In fact, the Irish mean score is not statistically different from a large number of Western European countries or from that of the USA[52]. Although, Ireland's mean score in mathematics is significantly below the OECD average, the score difference is only 9 points. While clearly there is room for improvement, this contrasts with other countries in the 'below-average' category, such as Greece (30 points below the OECD average) and Turkey (50 points below). Indeed, other country's scoring 6 points below the OECD average fall in the 'equivalent to the OECD average' category. In science, although Ireland is in the 'above-average' category, it exceeded the OECD average score by only 8 points. Ireland's mean science score is lower however than other 'above-average' countries, such as Finland (53 points above the OECD average) and New Zealand (31 points above it). Countries with scores in Science only somewhat less than Irelands fall in the 'equivalent to the OECD average' category. When the interpretation of what the magnitude of differences in score points actually mean unclear, some care should be

[52] As indicated both by score relative OECD mean and comparison country mean

taken in interpreting the result, in particular, where the poor rankings might be used to support significant policy change.

The PISA assessment demonstrates what can and is being achieved by 15 year olds (as captured by PISA) in other countries, thus providing a useful benchmark by which to consider the performance of Irish students. However, what is less clear is whether the data tells us something about the effectiveness of the Irish education system. While poor performance in PISA is often interpreted as indicating a deficiency in schools (in Ireland and elsewhere[53]), as a measure deemed to capture 'whole-life learning', it also reflects a range of other factors; country, school, classroom, family and individual effects[54]. As such, while a declining performance might, or might not, result from a reduction in the effectiveness of the education system; it also might result from a variety of other factors un-related to the education system, or indeed a combination of both[55]. It is worth noting, however, that in general, across countries, non-school factors (such as socio-economic status) are found to be associated with educational performance levels and one role of education systems, and schools, is to mitigate the educational impact of those differences (Lowell & Salzman, 2007: 22) (OECD, 2004b, 2010e)

PISA 2009 results suggest there has been some decline in relative student performance (as calculated by PISA) in Ireland in both reading and mathematics, with no change in science. The extent of the decline or the degree to which it reflects on the education system, schools or teachers is less clear. It could be that the effectiveness of Irish schools have declined over time thus impacting on student performance or it could be that some combination of a variety of other factors relating to school, classroom or family life are impacting. However, it could also be that the results from PISA are not giving a full account of the performance of Irish students. Identifying what factors are impacting positively or negatively is important because their effects may be enhanced or mitigated by public policy. Without alternative evidence, such as that which might be collected as part of a systematic evidenced based evaluation system, it is difficult to know. There is little or no data collected at post-primary level to provide corroborating evidence either way.

---

[53] (Lowell & Salzman, 2007)

[54] "If a country's scale scores in reading, scientific or mathematical literacy is higher than in another country, it cannot automatically be inferred that the schools or particular parts of the education systems in the first country are more effective than those in the second. However one can legitimately conclude that the cumulative impact of learning experiences in the first country, starting in early childhood and up to the age 15, embracing experiences both in school, home and beyond, have resulted in higher outcomes in the literacy domains that PISA measures" (OECD 2010a: 171)

[55] For example, one finding from PISA is that pre-primary care has a positive impact on student performance (OECD, 2010a: 6).

This leads to two main conclusions. First a nuanced and careful examination of PISA results is required and caution should be applied when extrapolating from them. Second, a multiplicity of data and information sources is needed to properly assess educational performance in Ireland. Without proper data and detailed analysis it is difficult to know how Irish schools are performing, how performance is changing over time and how continuous improvement can be supported through policy.

Interpreting Changes in PISA scores for Ireland

The Educational Research Centre (ERC) in Ireland has undertaken a variety of detailed studies on Ireland performance in PISA. This analysis suggests, with respect to the decline in both literacy and maths and the stability in performance in science, that a range of factors are implicated in influencing the performance in each domain(Perkins *et al.*, 2010, Cosgrove, J. *et al.*, 2010, Perkins *et al.*, 2011). As such, while the declines in reading performance between 2000 and 2009 and that of maths between 2003 and 2009 in PISA could indicate a real decline in the knowledge and skills of students[56], they could also be indicative of other factors associated with the test and its administration. Analysing the factors associated with the changes in Ireland performance, the ERC suggest there is some support for both explanations.

Research by the ERC both reviews and rules out a number of factors as grounds for providing any meaningful explanation for the changes in performance. Those reviewed and ruled out include sample design, achieved samples of school and students, the quality of national versions of the assessment instruments, and procedures used to administer the test (Perkins *et al.*, 2010, Cosgove & Hislop, 2011). The ERC is of the view that, aside from demographic changes in the PISA population, such as the increase in the percentage of immigrant student from 2.3% in 2000 to 8.3% in 2009, declines in the engagement of students with the PISA tests over time (as opposed to their ability to correctly respond to questions) have contributed to the decline in reported achievement scores (Perkins *et al.*, in press).However, it remains the view of the PISA Consortium (those responsible for the design and implementation of the surveys) that the declines in Irish performance in reading (since 2000) and maths (since 2003) indicate real declines in cognitive proficiency, and should be interpreted solely in the context of demographic and structural changes.

A number of such factors all considered relevant to an examination of performance change are highlighted(Perkins *et al.*, 2010);

[56] Independent analysis by Statistics Canada suggest that there has been a decline in achievement but that it is smaller in magnitude. (Cosgove & Hislop, 2011)

- *Demographic changes*: an increase in the proportion of students with an immigrant background/who speak a language other than English or Irish[57], a decrease in the proportion of early school leavers (a positive development in itself but one which may negatively impact on scores) and a decrease in the proportion of 15 years olds enrolled in a Leaving cert course[58].
- *Reading Habits:* a decrease in leisure reading, as well as lower reported enjoyment of reading for females
- *Chance Factors:* the chance sampling of 8 low performing schools[59]
- *Student Engagement:* evidence of less effort in 2009 than in previous assessments[60]
- *Method of Producing and Reporting Trends:* evidence that scaling and linking of data across cycles may have resulted in the reported results representing a over-estimate of the difference between 2000 and 2009 [61]

Characteristics Related to Achievement

Using PISA data, focusing on the factors associated with reading achievement only[62], the Educational Research Centre and the Irish Department of Education highlight some preliminary results from analysis based on multi-level modelling[63] and give an early indication of the importance of a number of variables related to achievement (Cosgrove, J. & Hislop, 2011, Perkins *et al.*, in press). A number of characteristics are shown to be related to achievement, these include School SSP status (DEIS)[64], immigrant/language status, parental occupation and education, early school leaving, engagement in reading activities, gender (and books at home), grade level, working part-time, absence from school and the use of meta-cognitive strategies.

---

[57] increase in the percentage of students with an immigrant background (2.3% in 2000 to 8.3% in 2009)

[58] Between 2000 and 2009 these was a marked decrease in the proportion of 15 years olds enrolled in a leaving cert course. The percentage in Transition year increased from 16% to 24%, this corresponded with a drop in the percentage in 5th year from 18.6% to 14.4%

[59] Although these schools have lower than average socio-economic status, after adjusting for student socio-economic status and gender, the schools still scored significantly below other schools, particularly in reading (Cosgove & Hislop, 2011: xi). It is suggested that the presence of these schools may be attributed to random sampling fluctuation or to some other factor or set of factors

[60] A decline a question answered in 4th half hour booklet.

[61] See (Perkins *et al.*, 2010)p57-63 and (Cosgove & Hislop, 2011)

[62] The score of the most recent major domain tends to be used as a proxy for overall student ability, performance in the three PISA domains are inter-related and tend to have similar relationships with explanatory variables

[63] in a recent presentation Cosgrove and Hislop 2011 these results will be published in 2012

[64] School Support Programme indicating disadvantaged school

Interestingly many of these factors could be classified as non-school factors, although schools do have a role in mitigating the effects of individual and family factors that might impact negatively on a student's performance. Other factors found to less relevant, all things being equal, include school sector, fee paying status, school location, school selectivity, use of ability and grouping, school climate, students family structure, as well as material and cultural possessions at home.

A recent paper by Smyth and McCoy (2011), considering the national and international literature on the characteristics related to education performance, summarises the evidence with respect to Ireland and identifies some of the challenges facing Irish second level education; these include inequality in educational outcomes related to socio-economic status, the proportion of young people entering second-level with low levels of literacy, a lack of student engagement with the teacher-centred methods used at second-level, limited catering within classrooms to the needs of range of ability levels, as well as a negative impact on the depth of learning experiences due to an exam focused approach to learning. With respect to student performance, the ESRI note that schools matter. This is true even where student characteristics are accounted for. Irish evidence, points to differences between schools across a number of student outcomes (achievement, attendance, early school leaving, subject take-up and personal development), regardless of student intake (This evidence is based mainly on analysis carried out by the ESRI using survey data and the ERC using PISA data[65]). Two factors are highlighted in particular as impacting on student performance. First, grouping students by ability is associated with greater inequality of outcomes leading to strong negative impacts for lower ability, without corresponding positive effects for those at the upper end. Second, school social climate, encompassing teacher-student relationships, is also highlighted as having an impact on student performance. The evidence on teacher effects is less clear, while international evidence suggests that both teachers and their approach to teaching can influence student performance, the authors point to a lack of systematic evidence on teaching methods at second level in Ireland.

Clearly, a systematic evidence based approach to educational evaluation supported by a comprehensive system of data collection and analysis would help further understanding of factors that support student performance and wider aspects of student development. It would also provide a strong evidence base for policy development and reform in the Irish education sector.

Exactly what form a systematic evidence-based approach to evaluation in education would take is not prescribed here. Nonetheless, it is important to note that it should

[65] See Smyth and McCoy, 2011, p7; Data used by the ESRI include the Co-education and Gender Equality Study in the 1990's and more recently the Post-Primary Longitudinal Study.

encompass a broad spectrum of data (from national-level, quantitative, systemic indicators to a systematic approach to the collection and analysis of contextual and qualitative data at the level of the school or classroom). Furthermore, it should be implemented in a way that supports systematic evidence-based evaluation, based on both summative and formative assessment, at both national and local level[66].

4 Conclusions

PISA is a useful systemic indicator of educational performance. The 2009 PISA assessment indicates that there has been some decline in Ireland's relative performance in Reading Literacy and Mathematics. While demographic change and other issues have been implicated in contributing to these observed declines, it is considered unlikely that they explain all of it. However, a cautious approach to the data and its interpretation is required. Focus on rank performance is misleading and fails to contextualise the results. A more nuanced consideration of the data suggests that while Ireland is not among the top performers in PISA, overall, Ireland's performance has been average, at the same level as many other western European nations and the US. Some caution should be applied in extrapolating from the results, in particular, where seemingly low rankings are used to support significant policy change without corroborating evidence. It is important to note, however, that advocating a more cautious and contextualised approach to the interpretation of PISA does not equate to claiming that the school system is performing well or that those with particular disadvantage are being well served[67].

PISA provides a useful comparator data source for the cross-country comparison of student performance, what can be achieved by students elsewhere. Considerable attention has been given both in Ireland and elsewhere to country rank performance in PISA. However overly focusing on these ranked league tables can lead to a simplistic interpretation of the results and have limited value from a policy perspective. While PISA is not without its limitations, it remains a useful and rich source of school, classroom, family and individual information from which lessons for policy and practice can be drawn. While PISA cannot establish causality, it can provide useful information on factors correlated with achievement. One general finding from the PISA assessment is that expenditure per student (resources) is not strongly associated with performance[68](OECD, 2010a: 160, 161). This coupled with

[66] Comparing similar systems with different outcomes, see McNamara et al for a comparison of the Irish and Icelandic approach to self-evaluation and their respective outcomes (McNamara *et al.*, 2011)

[67] This echo a point made by Lowell and Salzman (2007) in respect of US performance in PISA (Lowell & Salzman, 2007: 25).

[68] Analysis from PISA shows that overall there is a lack of relationship between resources and outcomes; this does not mean that resources are not important just that their level does not

the fact that expenditure per student in Ireland (in public institutions of primary, secondary and post-secondary non-tertiary education) rose considerably between 2000 and 2008 (83% in real terms[69]), suggests that securing improvements in educational performance are about more than resources[70]. Clearly, in a time of austerity, this is good news of a kind. PISA, in combination with other relevant evidence and information, can support informed judgements about the significance of reported outcomes (Cosgrove, J. & Hislop, 2011).

International and national education research provides the basis for a strong evidence-based approach to education policy[71]. In Ireland, the ERC provides detailed analysis of the PISA data. Yet, the OECD note that Ireland is among those countries who make limited use of the PISA data for decision-making, benchmarking and information purposes (OECD, 2011b, 2010f). This suggests more might be done to translate lessons from education research into policy development. Indeed, NESC, in a forthcoming report on standards and accountability in education in Ireland, points to the slow pace of policy development in education, more generally. While good policy development in education requires quality data and detailed analysis, something which is supported by a national standards infrastructure, it also requires a system wide culture and regime of evaluation coupled with the will and capacity to drive evidence-based change in both policy and practice.

Ireland took an uncritical approach to the PISA assessment when results suggested an average to above average performance (Eivers, 2010). The results from the 2009 assessment, which resulted in a more negative perception of student achievement in Ireland, has served to challenge both Ireland's use of PISA and indeed the system of education evaluation at national level. Three points emerge from this.

Firstly, performance aside, it is quite necessary and correct that there is a searching and detailed examination of what PISA does and does not tell us. PISA serves as a

---

have a systematic impact within the prevailing range. The OECD note that " if most or all schools have the minimum resource requirements to allow effective teaching, additional material resources may make little difference to the outcomes" (OECD, 2010b:14

69 (OECD, 2011b:120)

70 A recent paper by Smyth and McCoy (2011: 7) note that "there is a large body of research that show policy and practice at school level can make a substantive impact on student outcomes" and that many of these require only a "modest level of expenditure".

71 For example, while factors such as class-size receive considerable attention, the international evidence on student performance and class-size is mixed. In general, reductions in class-size are not considered the most cost-effective way of improving student attainment (Newman, 2011:369, Smyth, E. & McCoy, 2009). (UK Department of Education, 2011).

useful tool with which to consider and compare student performance but it must be used and interpreted with care.

Secondly, using PISA as the sole metric to validate Irish educational success, when the results are good, is just as unhelpful and inaccurate as using it to conclude that the 2009 PISA performance in Ireland was somehow catastrophic and that Ireland's education system is deficient. Regardless of whether relative student achievement in Ireland, as indicated by PISA, is high or low, PISA remains just one source of information on educational performance and should serve only as a complement to other international evidence and Ireland's own national system of assessment and evaluation. But this, of course, requires that Ireland's national standards infrastructure is both comprehensive and systematic.

Thirdly, overall, this serves to highlight the deficiencies in Ireland's approach to data gathering, analysis and evaluation in the Irish school sector. If PISA is not to remain a key indicator of Irish education performance then there must be a comprehensive and systematic national data and evaluation system through which the quality of teaching and learning can be monitored and improved. A systematic evidence based approach to evaluation in education should help close the informational deficit and deliver a multiplicity of evidence (quantitative and qualitative based on both summative and formative assessment) to support a broader perspective on Irish education.

The building of a national quality framework for education remains a challenge, as it is in other human services. This challenge is the subject of a NESC study of systems of quality, standards and accountability. The study describes the standards and quality system developed over the past decade in education, eldercare, end-of-life, disability and policing. Such detailed analysis provides the basis for the assessment of these systems and constructive discussion on how they might be improved. The paper on education, discussing Ireland approach to evaluation in education, provides a useful complement to this discussion of PISA.

BIBLIOGRAPHY

Cosgove, Jude & Hislop, Harold (2011), to.

Cosgrove, J., Shiel, G., Archer, P. & Perkins, R. (2010), *Comparisons of Performance in Ireland PISA 2000 to PISA 2009: A Preliminary Report to the Department of Education and Skills*, Educational Research Centre.

Cosgrove, J., Perkins, R., Moran, G. & Shiel, G. (2011), *Digital Reading Literacy in the OECD Programme for International Student Assessment (PISA 2009): Summary of Results for Ireland*, ERC.

Cosgrove, J. & Hislop, H. (2011), 'Looking back at PISA 2009 and forward to PISA 2012', Presentation to the National Assocation of Prinicipals and Deputy Principals Conference Killarney October 2011.

De Bartoli, L. & Thomson, S. (2010), *Contextual Factors that influence achievement of Australia's Indigeous students: Results from PISA 2000-2006*, ACER Press

Eivers, E. (2010), 'PISA: Issues in Implementation and Interpretation ', *The Irish Journal of Education,* 2010(xxxviii): 94-118.

Lowell, B. L. & Salzman, H. (2007), *Into the Eye of the Storm: Assessing the Evidence on Science and Engineering Education, Quality and Workforce Demand* Institute, The Urban.

McNamara, G. & O'Hara, J. (2006), 'Workable Compromise or Pointless Exercise?: School-based Evaluation in the Irish Context', *Educational Management Administration and Leadership,* 34: 564.

McNamara, G., O'Hara, J., Lisi, P. & Davidsdottir, S. (2011), 'Operationalising self-evaluation in schools: experiences from Ireland and Iceland', *Irish Educational Studies,* 30(1): 63-82.

NESC (2005), *NESC Strategy 2006: People, Productivity and Purpose*, No.114, December 2005, NESDO.

NESC (Forthcoming), *Quality and Standards in Human Services in Ireland: Education*, National Economic and Social Council.

Newman, C. (2011), 'Education: Market Failure and Government Interventions' in O'Hagan, J., Newman, C. (Ed.) *The Economy of Ireland* Gill and MacMillan

OECD (2004a), *Learning for Tomorrows World - First Results from PISA* OECD.

OECD (2004b), *Messages from PISA 2000*, OECD, http://www.oecd.org/dataoecd/31/19/34107978.pdf.

OECD (2005), *PISA 2003 Data Analysis Manual: SPSS Users*, OECD.

OECD (2007), *Improving School Leadership - Ireland*, Leadership, Prepared for the Irish Department of Education and Science for the OECD Activity Improving School, OECD, Leadership Development for Schools

OECD (2010a), *PISA 2009 Results: What Students Know and Can Do*, OECD.

OECD (2010b), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices, Volume IV*, OECD.

OECD (2010c), *PISA 2009 Results: Learning Trends, Changes in Student Performance since 2000, Volume V*, OECD.

OECD (2010d), *Tax Policy Reform and Economic Growth*, OECD Tax Policy Studies Report No. 20.
OECD (2010e), 'Socioeconomic'.
OECD (2010f), *What Makes Schools Successful?*
OECD (2011a), *OECD Review of Evaluation and Assessment in Education -Australia*, OECD.
OECD (2011b), *OECD Economic Survey for Ireland*.
Perkins, R., Moran, G., Cosgrove, J. & Shiel, G. (2010), *PISA 2009: The Performance and Progress of 15 year olds in Ireland: Summary Report*, Educational Research Centre.
Perkins, R., Moran, G., Cosgove, J. & Shiel, G. (2011), *PISA 2009:The Performance and Progress of 15 year olds in Ireland: Submission to the Joint Oireachtas Committee on Education and Skills*, Educational Research Centre.
Perkins, R., Cosgrove, J., Moran, G. & Shiel, G. (in press), *PISA 2009: The performance and progress of 15-year-olds in Ireland – Main report*, Dublin: Educational Research Centre.
Sabel, C., Saxenian, A., Miettinen, R., Kristensen, P. H. & Hautamäki, J. (2011), *Individualized service provision in the new welfare state: lessons from special education in Finland.* , Helsinki: Sitra. .
Shiel, G., Kellaghan, T. & Moran, G. (2010), *Standardised Testing in Lower Secondary Education*, 12, NCCA Report No., Educational Research Centre.
Smyth, E. & McCoy, S. (2009), *Investing in Education*, ESRI.
Smyth, E. & McCoy, S. (2011), 'Improving Second-Level Education: Using Evidence for Policy Development ', *ESRI Renewal Series,* Paper 5.
UK Department of Education (2011), *Class-room Size and Education in England: Evidence Report*, Research report DFE-RR 169.