# PIRLS & TIMSS 2011

## Reading, Mathematics and Science Outcomes for Ireland

Eemer Eivers and Aidan Clerkin

Educational Research Centre

# PIRLS & TIMSS 2011

## Reading, Mathematics and Science Outcomes for Ireland

Eemer Eivers and Aidan Clerkin

Educational Research Centre

Cataloguing-in-Publication Data

Eivers, Eemer

PIRLS and TIMSS 2011: Reading, Mathematics and Science Outcomes for Ireland / Eemer Eivers and Aidan Clerkin. Dublin: Educational Research Centre.

iv, 44 p.; 30cm
Includes bibliographical references.

1. Reading (Primary) – Ireland – Evaluation
2. Mathematics (Primary) – Ireland – Evaluation
3. Science (Primary) – Ireland – Evaluation
4. Academic achievement
5. Educational tests and measurements

2012
I Clerkin, Aidan
371.262-dc23

ISBN: 978-0-900440-40-3

Design by SilverBark Creative
Printed in the Republic of Ireland by ePrint Limited

# Contents

# Preface

PIRLS (Progress in International Reading Literacy Study) and TIMSS (Trends in International Mathematics and Science Study) are projects of the International Association for the Evaluation of Educational Achievement (IEA). They are designed to assess the reading, mathematics and science achievement of Fourth class pupils. TIMSS was first conducted in 1995 while PIRLS first took place in 2001. In 2011, Ireland took part in PIRLS for the first time, and in TIMSS for the first time since 1995.

The present volume is published at the same time as the IEA's three main reports on the three domains of reading (Mullis, Martin, Foy, & Drucker, 2012b), mathematics (Mullis, Martin, Foy, & Arora, 2012a), and science (Martin, Mullis, Foy, & Stanco, 2012). It summarises the main achievement-related findings from both studies, focussing on Irish performance and findings most likely to be of interest to an Irish audience. A more in-depth series of reports on PIRLS and TIMSS 2011 from an Irish perspective will be released in early 2013. Later in 2013, the IEA will publish a report examining the relationship between performance on all three domains.

## Acknowledgements

We acknowledge the support of the National Advisory Committee which was appointed by the Minster of Education and Skills to assist in implementing PIRLS and TIMSS. The members of the Committee are: Carmel Nic Airt (An Foras Pátrúnachta), Dónal Ó hAiniféin (Gaelscoileanna) (replaced by Máirín Ní Chéileachair from May 2012), Mark Candon (CPSMA), Deirbhile Nic Craith (INTO), John Curran (IPPN) (replaced by Aedín Ní Thuathail from October 2011), Arlene Forster (NCCA), Áine Lynch (NPC–P), and Ciara O'Donnell (PDST) (replaced by Mary Manley in October 2011). The Department of Education and Skills is represented by Margaret Kelly (replaced by Breda Naughton from May 2012), Pádraig Mac Fhlannchadha (Chair until July 2011) and Caitríona Ní Bhriain (Chair from August 2011).

Thanks are also due to staff at the ERC, including Gerry Shiel and Peter Archer (Director), each of whom acted as a National Research Coordinator for the field trial, and to Mary Lewis, who managed the PIRLS scoring. Many thanks to Joanne Kiniry, Rose Fish and Ann-Marie Creaven, who worked as research assistants on PIRLS and TIMSS, to Seán Close, who reviewed parts of this report, and to David Millar, who advised on sampling. Thanks also to Mary Rohan, Paula Chute, John Coyle, Caoimhe Máirtín, and Hilary Walshe for their assistance and support. Finally, and most importantly, we thank the school principals, teachers, pupils and parents who participated in the study.

# 1 Introduction

In 2011, Ireland participated in two major international studies assessing the skills of Fourth class pupils. PIRLS (**P**rogress in **I**nternational **R**eading **L**iteracy **S**tudy) examined pupils' reading achievement, while TIMSS (**T**rends in **I**nternational **M**athematics and **S**cience **S**tudy) examined mathematics and science achievement. This report summarises the main findings of the studies, focussing specifically on Ireland's performance and factors related to achievement among Irish pupils. It can be read in conjunction with the three main international reports and the encyclopaediae for the two studies, all of which are available from http://timssandpirls.bc.edu/.

The international reports for reading (Mullis, Martin, Foy & Drucker, 2012b), mathematics (Mullis, Martin, Foy & Arora, 2012a) and science (Martin, Mullis, Foy, & Stanco, 2012) provide a broad description of the performance of pupils in all participating countries, and relate performance to selected characteristics of individual pupils, and their home, class and school environment. The encyclopaediae for PIRLS (Mullis, Martin, Minnich, Drucker & Ragan, 2012c) and TIMSS (Mullis et al., 2012d) provide a context within which to understand the results. For example, they can be used to compare curriculum coverage for the target grades in each of the participating countries. More detailed *national* analyses of Irish performance on PIRLS and TIMSS, including examples of the types of test items on which Irish pupils excelled or struggled, will follow in early 2013. Also, *international* analyses on differential school effectiveness and on the relationship between performance on reading, mathematics and science will be published later in 2013.

The present volume is divided into seven chapters. This chapter provides a short introduction to the studies, explaining what they assess, how the assessment is conducted, and how to interpret the achievement data presented. Chapter 2 presents the main reading, mathematics and science results for *all* participating countries, and outlines any statistically significant gender differences. Chapter 3 describes performance on International Benchmarks for reading, mathematics and science. Chapters 4, 5 and 6 present more detailed analyses of Irish performance on reading, mathematics and science, respectively. They include a description of performance on the respective subscales, nationally and by gender. In chapters 3 to 6, Irish performance is compared with a subset of key comparison countries (the top-performing countries on a domain, plus English-speaking countries), and information related to performance trends and areas of relative strength and weakness is discussed. Chapter 7 summarises the findings, notes future directions, and outlines some forthcoming reports.

## What are PIRLS and TIMSS?

PIRLS and TIMSS are large international comparative studies of achievement that assess, respectively, the reading, and mathematics and science skills of primary school pupils. TIMSS, but not PIRLS, also has a post-primary component. The studies are conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), and are managed at an international level by the International Study Center in Boston College. Within each participating country, a National Research Centre – sometimes the Education Ministry – manages the study or studies. In Ireland, this role was filled by the Educational Research Centre.

First conducted in 1995, TIMSS takes place every four years, assessing the mathematical and scientific skills of pupils in Fourth and/or Eighth grade (equivalent to Fourth class and Second Year in Ireland). In contrast, PIRLS takes place every five years, assesses Fourth grade only, and was first conducted in 2001. In 2011, the cycles for the studies coincided for the first time, giving countries the opportunity to take part in one or both assessments. Sixty-three countries took part in TIMSS (at either Fourth or Eighth grade, or both), 49 took part in PIRLS, and 34 took part in both assessments at Fourth grade. Table 1.1 lists all participating countries and benchmarking participants[1] (e.g., Quebec in Canada, Andalusia in Spain). Across all participating countries, the tests were administered in a total of 58 languages, with English and Arabic being the two most common languages for both TIMSS and PIRLS.

At the Fourth grade alone, a total of almost 300,000 pupils took part in PIRLS and 290,000 in TIMSS. There was considerable overlap between the two studies, and a majority of Fourth grade participants took part in both.

**Table 1.1: Countries and benchmark participants that took part in the 2011 cycle of PIRLS, TIMSS or both**

<table>
<tr><th colspan="2">PIRLS & TIMSS</th><th>TIMSS only</th><th>PIRLS only</th></tr>
<tr><td>Australia</td><td>New Zealand</td><td>Armenia</td><td>Belgium (French)</td></tr>
<tr><td>Austria</td><td>Northern Ireland</td><td>Bahrain</td><td>Bulgaria</td></tr>
<tr><td>Azerbaijan</td><td>Norway</td><td>Belgium (Flemish)</td><td>Canada</td></tr>
<tr><td>Botswana</td><td>Oman</td><td>Chile</td><td>Colombia</td></tr>
<tr><td>Chinese Taipei</td><td>Poland</td><td>Ghana</td><td>France</td></tr>
<tr><td>Croatia</td><td>Portugal</td><td>Japan</td><td>Trinidad and Tobago</td></tr>
<tr><td>Czech Republic</td><td>Qatar</td><td>Jordan</td><td rowspan="19">Benchmark<br>Andalusia (Spain)<br>South Africa (Eng./ Afrikaans)<br>Malta (Maltese)</td></tr>
<tr><td>Denmark</td><td>Romania</td><td>Kazakhstan</td></tr>
<tr><td>England</td><td>Russian Fed.</td><td>Korea, Republic of</td></tr>
<tr><td>Finland</td><td>Saudi Arabia</td><td>Lebanon</td></tr>
<tr><td>Georgia</td><td>Singapore</td><td>Macedonia</td></tr>
<tr><td>Germany</td><td>Slovak Republic</td><td>Malaysia</td></tr>
<tr><td>Honduras</td><td>Slovenia</td><td>Palestinian Nat'l Auth.</td></tr>
<tr><td>Hong Kong SAR</td><td>Spain</td><td>Serbia</td></tr>
<tr><td>Hungary</td><td>Sweden</td><td>South Africa</td></tr>
<tr><td>Indonesia</td><td>United Arab Emirates</td><td>Syria</td></tr>
<tr><td>Iran</td><td>United States</td><td>Thailand</td></tr>
<tr><td>Ireland</td><td rowspan="8">Benchmark<br>Abu Dhabi & Dubai (UAE)<br>Alberta, Ontario & Quebec (Canada), Florida (USA)</td><td>Tunisia</td></tr>
<tr><td>Israel</td><td>Turkey</td></tr>
<tr><td>Italy</td><td>Ukraine</td></tr>
<tr><td>Kuwait</td><td>Yemen</td></tr>
<tr><td>Lithuania</td><td rowspan="4">Benchmark<br>Alabama, California, Colorado, Connecticut, Indiana, Mass., Minnesota & North Carolina (USA)</td></tr>
<tr><td>Malta</td></tr>
<tr><td>Morocco</td></tr>
<tr><td>Netherlands</td></tr>
</table>

[1] Benchmarking participants are regional entities that follow the quality standards established by TIMSS/PIRLS. Their data are comparable to the countries' data, and they can use the results as a benchmark.

To accommodate developing countries, the option to assess pupils at a higher grade or to participate in prePIRLS (a less difficult version of the PIRLS assessment) was provided. Information on the small number of countries who used out-of-grade assessment or took part in prePIRLS is reported in the main international reports, but not included here.

## What do PIRLS and TIMSS assess?

Test content in PIRLS and TIMSS is guided by assessment frameworks, which are briefly summarised here. The full PIRLS (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009) and TIMSS (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009) frameworks can be found at http://timssandpirls.bc.edu/.

### PIRLS assessment framework

The PIRLS assessment is built around two organising dimensions: reading purpose and comprehension process. Purpose refers to *why* readers read a text – either for literary experience or to acquire and use information. Comprehension processes examine *how* readers process what they read. PIRLS measures four comprehension processes: focussing on and retrieving explicitly stated information; making straightforward inferences; interpreting and integrating ideas and information; and, examining and evaluating content, language and textual elements.

The PIRLS assessment materials used in the 2011 cycle consisted of 10 discrete texts, five of which were Literary and five Informational. Each text had a number of related items assessing a mixture of comprehension processes. Table 1.2 summarises the percentages of the test allocated to each comprehension process, as specified by the PIRLS framework. Overall, 30% of the test items assessed making Straightforward Inferences and 30% assessed Interpreting and Integrating. Retrieving Information and Examining and Evaluating content were each represented by 20% of items (Table 1.2).

**Table 1.2: Percentages of the PIRLS reading assessments devoted to reading processes**

| Reading Process | % of the test |
|---|---|
| Focus on & retrieve explicitly stated information | 20 |
| Make straightforward inferences | 30 |
| Interpret & integrate ideas & information | 30 |
| Examine & evaluate content, language & textual elements | 20 |

### TIMSS assessment framework

The assessment of mathematics and science in TIMSS is also built around two organising dimensions: content and cognition. Content refers to the subject matter to be assessed while cognition deals with the thinking processes expected of pupils as they engage with the content. The three content and three cognitive domains that were assessed for mathematics and science are summarised in Table 1.3. Also shown are the percentages of the test allocated to each dimension, as specified by the TIMSS framework. For both mathematics and science, the cognitive processes that were targeted are Knowing, Applying, and Reasoning. Knowing and Applying were each allocated 40% of test content, with the remaining 20% devoted to Reasoning. The mathematics framework allocated half the assessment to the content area of Number, 35% to Geometric Shapes and Measures, and the remainder to Data Display. For science, almost half of the testing framework involved Life science, 35% related to Physical science and 20% related to Earth science.

| Table 1.3: Target percentages of the TIMSS mathematics and science assessments devoted to content and cognitive domains at Fourth grade | | | | | |
|---|---|---|---|---|---|
| Content | | | | Cognitive Processes | |
| Maths | | Science | | Maths & Science | |
| Topic | % | Topic | % | Process | % |
| Number | 50 | Life Science | 45 | Knowing | 40 |
| Geometric Shapes & Measures | 35 | Physical Science | 35 | Applying | 40 |
| Data Display | 15 | Earth Science | 20 | Reasoning | 20 |

### Contextual information

As well as gathering achievement data, both studies used questionnaires to gather contextual information from pupils, their teachers and school principals. PIRLS – but not TIMSS – also collected information from parents. In PIRLS and TIMSS 2011, over a quarter of a million parents answered questionnaires; about 14,000 teachers completed questionnaires for PIRLS and/or TIMSS, as did approximately 10,000 principals (Table 1.4).

| Table 1.4: Number of completed questionnaires for all participants, PIRLS and TIMSS 2011 | | | |
|---|---|---|---|
| | Parents | Teachers | Schools |
| PIRLS/prePIRLS | 263,308 | 13,998 | 10,297 |
| TIMSS 4th grade | n/a | 13,527 | 9,783 |

In addition, each participating country submitted a chapter for the PIRLS encyclopaedia (Mullis et al., 2012c), a chapter for the TIMSS encyclopaedia (Mullis et al., 2012d) and completed a detailed questionnaire about the national curricula for reading, mathematics and science. These data provided national contextual information about issues such as structure of the education system, teacher professional development, and curriculum. Thus, the studies have a very rich backdrop against which to interpret the achievement data.

## Structure of TIMSS and PIRLS assessments

TIMSS and PIRLS are large-scale assessments of core curriculum areas, both requiring administration of a large set of test items. As such, each pupil saw only a subset of the total number of test items (questions). Test items were grouped into blocks of items. Each block appeared in different positions and combinations in booklets, allowing responses to be linked together across booklets. There were 14 different TIMSS and 13 different PIRLS test booklets. Each pupil completed one PIRLS and one TIMSS test booklet, randomly assigned to ensure that similar proportions of pupils responded to each booklet. In each TIMSS booklet, one half was devoted to science and one half to mathematics.

### Question types and scoring

Two question formats were used in TIMSS and PIRLS – multiple-choice and constructed-response. Multiple-choice questions have four possible response options, only one of which is correct. Constructed-response questions require a written response, which can range in length from a single word or number to a paragraph. Each constructed-response item has an accompanying scoring guide that details how to score pupils' responses. Some constructed-response items were full credit (i.e., the response is either right or wrong) while others were partial credit items (e.g., a fully correct answer gets 2 points, a partially correct answer gets 1 point, and an incorrect answer gets 0). When scoring, emphasis was placed on the quality of the answer with respect to the topic in question.

Writing or spelling ability was not considered, once answers were communicated clearly enough to be understood by scorers.

### Study design and quality control

In all participating countries, similar sampling techniques were used, and sampling was independently monitored by Statistics Canada to ensure comparability. All participating countries had to reach at least 95% coverage of the target population (e.g., Fourth class pupils) in their sampling frame and to achieve school- and pupil-level participation rates of at least 85%. Additional quality control mechanisms at the pre-testing, testing and post-testing phases of the studies ensured a high degree of accuracy in the data gathered and meant that cross-country comparisons could be supported. Quality control measures used included:

- independent verification of all national translations and adaptations of instruments
- national and international quality monitoring programmes, each visiting 10% of participating schools on the days the tests were administered
- multiple-marking of approximately 25% of constructed-response items in each country, to assess scorer reliability.

See http://timssandpirls.bc.edu/methods/index.html for details of PIRLS and TIMSS methods and procedures (including sampling, translation, and quality assurance). Details of how these procedures were implemented in Ireland are available in the national technical report (Eivers & Clerkin, 2012).

## PIRLS and TIMSS in Ireland

In 2011, Ireland took part in both PIRLS and TIMSS (at Fourth class only). This was the first time that Ireland had taken part in TIMSS since 1995, when both Third and Fourth class pupils were assessed. Then, Irish Fourth class pupils ranked sixth of 17 countries for mathematics (Mullis, Martin, Beaton, Gonzalez, Kelly & Smith, 1997), and ranked eighth of 17 countries for science (Martin, Mullis, Beaton, Gonzalez, Smith & Kelly, 1997). For both domains, the mean for Fourth class pupils was significantly higher than the international averages.

Ireland had not participated in any of the previous cycles of PIRLS, but did participate in the IEA's 1991 Reading Literacy Study – widely regarded as the precursor to PIRLS. In that study, Irish 9-year-olds scored close to the overall average, ranking 12th of 27 participating countries (Martin & Morgan, 1994).

Ireland's limited participation in large international studies at *primary* level is in contrast with recent national participation in a number of post-primary level studies such as PISA (Programme for International Student Assessment), ICCS (International Civic and Citizenship Education Study) and TALIS (Teaching and Learning International Survey). Recognition of the gap in comparative information at primary level led to the Department of Education and Skills' decision in late 2009 to participate in an international assessment at primary level. In addition, the fact that the PIRLS and TIMSS study cycles would coincide meant that achievement data across all three domains could be related. As a result, Ireland took part in the PIRLS and TIMSS field trial in 2010 and main study in 2011. Irish participation in PIRLS and TIMSS was managed by the Educational Research Centre on behalf of the Department of Education and Skills. An advisory committee, composed of the main education partners, oversaw the implementation of the studies and assisted with the interpretation of outcomes.

## Who took part?

A list was generated of all schools with Fourth class pupils in Ireland, and 152 schools were randomly selected (balanced by enrolment size, language of instruction, DEIS status and gender mix) from the list. Within each selected school, up to two Fourth class groups were randomly chosen to participate. In schools with fewer than three Fourth class groups, all were selected. Of the 152 selected schools, one had closed. A further three were unable to take part, and were replaced by three pre-assigned replacement schools with similar characteristics. Within the final sample of 151 participating schools, 4825 Fourth class pupils were selected to participate.

Forty-three pupils (less than 1%) were excluded from both assessments, typically because of an intellectual disability or limited English proficiency. In total, 4560 pupils completed the TIMSS assessment and 4524 completed PIRLS (Table 1.5). Thus, with an initial school-level participation rate of 98.0% (100% with replacements) and weighted pupil participation rates of 95%, Ireland comfortably exceeded the required minimum school- and pupil-level participation rates[2]. Ireland also had excellent response rates to all questionnaires – ranging from 93.8% for the PIRLS test and the parental *Learning to Read Survey* to 99.5% for the Teacher Questionnaire. The high levels of cooperation from pupils, parents and school staff, as reflected in the very high participation and response rates, mean that the data are likely to be an accurate reflection of the achievements, attitudes and environment of Fourth class pupils.

**Table 1.5: Response rates to tests and questionnaires in Ireland, PIRLS and TIMSS 2011**

| Instrument | No. of pupils = 4825 | |
|---|---|---|
| | N | % |
| TIMSS test | 4560 | 94.5 |
| PIRLS test | 4524 | 93.8 |
| Pupil questionnaire | 4568 | 94.7 |
| Parent questionnaire | 4524 | 93.8 |
| | No. of classes = 221 | |
| Teacher questionnaire | 220 | 99.5 |
| | No. of schools = 151 | |
| School questionnaire | 145 | 96.0 |

Participating pupils in Ireland had an average age of 10.3 years, the same as the international average ages for both PIRLS and TIMSS. The average testing age in almost all countries fell between 9.9 and 10.7 years, with most clustered close to the study average. Participants in Ireland were almost evenly divided by gender (51.1% boys and 48.9% girls), and a significant minority were enrolled in schools where the medium of instruction was Irish (7.7% for PIRLS and 7.4% for TIMSS participants). Schools teaching through Irish had the option of administering the TIMSS test in Irish,[3] and five of the 10 schools in question (3.3% of pupils) did so. Full details of all sampling procedures, participation and response rates in Ireland are provided in Ireland's national technical report (Eivers & Clerkin, 2012).

[2] Due to errors in their administration of the TIMSS test, data for one Irish school are included in the PIRLS, but not the TIMSS, dataset.

[3] In Ireland, PIRLS was considered to be a test of *English* reading, and was administered in English in all cases.

## Reader guide

### Centrepoints

PIRLS and TIMSS compare country performance against a scale centrepoint of 500. The centrepoint is a point of reference that remains constant from assessment to assessment. It is the mean score from the first time the study was conducted. Thus, the overall reading scale centrepoint relates to 2001, and the mathematics and science overall scales to 1995. In contrast, the international average of the mean scores for each of the participating countries changes from cycle to cycle. In 2011, for example, the international average is below 500 for mathematics and science, and above 500 for reading.

### Rounding

Numbers in some tables may not add up exactly to the totals, due to rounding. All totals, differences between scores and averages are calculated using exact numbers, which are rounded only after calculation. Thus, the points difference between two scores as described in the text may differ marginally from the difference between the rounded scores shown in a table.

### Scale scores

PIRLS and TIMSS use item response theory methods to describe achievement on a scale with an average of 500 and a standard deviation of 100. This means that 68% of pupils' scores fall between 400 and 600 (the average score ± one standard deviation). The scales for reading, mathematics and science use the same general structure, but are not directly comparable. For example, we cannot say that a mathematics score of 550 and a science score of 550 means that a pupil has equal amounts of mathematical and scientific "knowledge". Comparisons should only be made within scales.

### Standard errors

Test scores obtained by groups of pupils are used to estimate the mean score for country populations. Thus, as well as the mean, we report the standard error. It is an estimate of how accurately the sample mean reflects the population mean, with smaller standard errors indicating a more precise estimate.

### Subscale and overall scale scores

Although one might expect a direct link between subscale scores and the corresponding overall score, there is no such algebraic equivalence. Overall scales are created independently of the subscales. What distinguishes them is whether the items are pooled into sub-groups or treated as a single assessment. Thus, subscale and overall scale scores will always be in close, but not always exact, agreement.

## When is a difference *statistically significant*?

Throughout this report, when we write about a *significant* difference, it refers to a difference between groups that a statistical test has established is unlikely to be due to chance.

As a rule of thumb, a difference of two or three points between the means of two countries is not meaningful. To check if a difference between two mean scores is significant or not, you have to consider two pieces of information:

- the points difference
- the size of the standard errors (shown as *SE* in tables).

### Testing for significance

The method used by PIRLS and TIMSS to test for significance is explained below. As it is a quite detailed calculation, we also provide a simple "back-of-the-envelope" calculation that is reasonably accurate.

#### Rough test

For each mean score, multiply its standard error by 2 to create a band around the mean score. If one band does not overlap with another, the difference may be significant. Take two means of 500 (SE=2.5) and 510 (SE=3.0). The band for the first mean is 495-505 (i.e., 500 ± 5) and for the second is 504-516. Because they overlap, the 10-point gap between the two mean scores is not significant. NOTE: This method slightly under-identifies significant differences and should only be used as a rough estimate.

#### Precise test

Differences in mean achievement between countries are considered statistically significant if the absolute difference between them, divided by the standard error of the difference, is greater than the critical value. For differences between countries, which can be considered as independent samples, the standard error of the difference between means is computed as the square root of the sum of the squared standard errors of each mean:

$$se_{diff} = \sqrt{se_1^2 + se_2^2}$$

where $se_1$ and $se_2$ are the standard errors of the means. Significance is not adjusted for multiple comparisons among countries (e.g., Bonferroni adjustment). Although such adjustments guard against misinterpreting the outcome of multiple simultaneous significance tests, the results vary depending on the number of countries included in the adjustment, leading to apparently conflicting results from comparisons using different combinations of countries.
[Adapted from the PIRLS 2006 Technical Report (Martin, Mullis & Kennedy, 2007)]

# 2 Overall performance on reading, mathematics and science

Each of the three domains (reading, mathematics and science) is scaled to have a study centrepoint score of 500 and a standard deviation of 100. As noted in Chapter 1, the centrepoint is the *anchor mean* from when the study was first conducted (2001 for PIRLS and 1995 for TIMSS). The international averages for 2011 are slightly below 500 for mathematics and science, and slightly above 500 for reading.

Tables 2.1 and 2.2 summarise the mean scores and position relative to Ireland, for each participating country and for benchmarking participants. Full details of the relative performance of all participating countries, including 95% confidence intervals, are available in the PIRLS and TIMSS international reports (Martin et al., 2012; Mullis et al., 2012a, 2012b). In early 2013, a set of thematic reports for Ireland (www.erc.ie/pirlstimss.ie) will provide additional details of Irish performance on reading, mathematics and science.

## What is shown in Tables 2.1 and 2.2?

The tables show the mean scores and standard errors for participating countries (Table 2.1) and benchmarking participants (Table 2.2) on reading, mathematics and science. Within each domain, countries are sorted in descending order by their mean scores. The tables compare each country mean with the study centrepoint and with Ireland's mean in 2011.

Green shading indicates a mean significantly higher than the scale centrepoint, red indicates a mean significantly lower than the centrepoint, and unshaded indicates no significant difference.

The IRL columns compare each country's mean to Ireland's. For a given country, the symbol ▲ denotes a significantly higher mean score, ◊ denotes a mean score that does not differ significantly from Ireland's, and ▼ denotes a significantly lower mean score.

See the box at the end of Chapter 1 for more information on what is meant by a significant difference.

### Reading

The highest performing countries on the reading test were Hong Kong, the Russian Federation, Finland and Singapore. Pupils in these four countries significantly outperformed pupils in *all* other participating countries. The mean scores for pupils in the top four countries ranged from 567 to 571 and were at least two-thirds of a standard deviation above the PIRLS centrepoint.

## Table 2.1: Mean country scores and standard errors for each domain in PIRLS and TIMSS 2011, and position relative to the study centrepoints and Irish means

| Reading | Mean | *SE* | IRL |
|---|---|---|---|
| Hong Kong | 571 | *2.3* | ▲ |
| Russian Fed. | 568 | *2.7* | ▲ |
| Finland | 568 | *1.9* | ▲ |
| Singapore | 567 | *3.3* | ▲ |
| N. Ireland | 558 | *2.4* | ▲ |
| United States | 556 | *1.5* | ◆ |
| Denmark | 554 | *1.7* | ◆ |
| Croatia | 553 | *1.9* | ◆ |
| Chinese Taipei | 553 | *1.9* | ◆ |
| **Ireland** | **552** | *2.3* | |
| England | 552 | *2.6* | ◆ |
| Canada | 548 | *1.6* | ◆ |
| Netherlands | 546 | *1.9* | ◆ |
| Czech Rep. | 545 | *2.2* | ◆ |
| Sweden | 542 | *2.1* | ▼ |
| Italy | 541 | *2.2* | ▼ |
| Germany | 541 | *2.2* | ▼ |
| Israel | 541 | *2.7* | ▼ |
| Portugal | 541 | *2.6* | ▼ |
| Hungary | 539 | *2.9* | ▼ |
| Slovak Rep. | 535 | *2.8* | ▼ |
| Bulgaria | 532 | *4.1* | ▼ |
| New Zealand | 531 | *1.9* | ▼ |
| Slovenia | 530 | *2.0* | ▼ |
| Austria | 529 | *2.0* | ▼ |
| Lithuania | 528 | *2.0* | ▼ |
| Australia | 527 | *2.2* | ▼ |
| Poland | 526 | *2.1* | ▼ |
| France | 520 | *2.6* | ▼ |
| Spain | 513 | *2.3* | ▼ |
| Norway | 507 | *1.9* | ▼ |
| Belgium (Fr.) | 506 | *2.9* | ▼ |
| Romania | 502 | *4.3* | ▼ |
| **Centrepoint** | **500** | – | ▼ |
| Georgia | 488 | *3.1* | ▼ |
| Malta | 477 | *1.4* | ▼ |
| Trinidad & Tobago | 471 | *3.8* | ▼ |
| Azerbaijan | 462 | *3.3* | ▼ |
| Iran | 457 | *2.8* | ▼ |
| Colombia | 448 | *4.1* | ▼ |
| UAE | 439 | *2.2* | ▼ |
| Saudi Arabia | 430 | *4.4* | ▼ |
| Indonesia | 428 | *4.2* | ▼ |
| Qatar | 425 | *3.5* | ▼ |
| Oman | 391 | *2.8* | ▼ |
| Morocco | 310 | *3.9* | ▼ |

| Maths | Mean | *SE* | IRL |
|---|---|---|---|
| Singapore | 606 | *3.2* | ▲ |
| Korea, Rep. | 605 | *1.9* | ▲ |
| Hong Kong | 602 | *3.4* | ▲ |
| Ch. Taipei | 591 | *2.0* | ▲ |
| Japan | 585 | *1.7* | ▲ |
| N. Ireland | 562 | *2.9* | ▲ |
| Belgium (Fl.) | 549 | *1.9* | ▲ |
| Finland | 545 | *2.3* | ▲ |
| England | 542 | *3.5* | ▲ |
| Russian Fed. | 542 | *3.7* | ▲ |
| United States | 541 | *1.8* | ▲ |
| Netherlands | 540 | *1.7* | ▲ |
| Denmark | 537 | *2.6* | ▲ |
| Lithuania | 534 | *2.4* | ◆ |
| Portugal | 532 | *3.4* | ◆ |
| Germany | 528 | *2.2* | ◆ |
| **Ireland** | **527** | *2.6* | |
| Serbia | 516 | *3.0* | ▼ |
| Australia | 516 | *2.9* | ▼ |
| Hungary | 515 | *3.4* | ▼ |
| Slovenia | 513 | *2.2* | ▼ |
| Czech Rep. | 511 | *2.4* | ▼ |
| Austria | 508 | *2.6* | ▼ |
| Italy | 508 | *2.6* | ▼ |
| Slovak Rep. | 507 | *3.8* | ▼ |
| Sweden | 504 | *2.0* | ▼ |
| Kazakhstan | 501 | *4.5* | ▼ |
| **Centrepoint** | **500** | – | ▼ |
| Malta | 496 | *1.3* | ▼ |
| Norway | 495 | *2.8* | ▼ |
| Croatia | 490 | *1.9* | ▼ |
| New Zealand | 486 | *2.6* | ▼ |
| Spain | 482 | *2.9* | ▼ |
| Romania | 482 | *5.8* | ▼ |
| Poland | 481 | *2.2* | ▼ |
| Turkey | 469 | *4.7* | ▼ |
| Azerbaijan | 463 | *5.8* | ▼ |
| Chile | 462 | *2.3* | ▼ |
| Thailand | 458 | *4.8* | ▼ |
| Armenia | 452 | *3.5* | ▼ |
| Georgia | 450 | *3.7* | ▼ |
| Bahrain | 436 | *3.3* | ▼ |
| UAE | 434 | *2.0* | ▼ |
| Iran, Rep. | 431 | *3.5* | ▼ |
| Qatar | 413 | *3.5* | ▼ |
| Saudi Arabia | 410 | *5.3* | ▼ |
| Oman | 385 | *2.9* | ▼ |
| Tunisia | 359 | *3.9* | ▼ |
| Kuwait | 342 | *3.4* | ▼ |
| Morocco | 335 | *4.0* | ▼ |
| Yemen | 248 | *6.0* | ▼ |

| Science | Mean | *SE* | IRL |
|---|---|---|---|
| Korea, Rep. | 587 | *2.0* | ▲ |
| Singapore | 583 | *3.4* | ▲ |
| Finland | 570 | *2.6* | ▲ |
| Japan | 559 | *1.9* | ▲ |
| Russian Fed. | 552 | *3.5* | ▲ |
| Ch. Taipei | 552 | *2.2* | ▲ |
| United States | 544 | *2.1* | ▲ |
| Czech Rep. | 536 | *2.5* | ▲ |
| Hong Kong | 535 | *3.8* | ▲ |
| Hungary | 534 | *3.7* | ▲ |
| Sweden | 533 | *2.7* | ▲ |
| Slovak Rep. | 532 | *3.8* | ▲ |
| Austria | 532 | *2.8* | ▲ |
| Netherlands | 531 | *2.2* | ▲ |
| England | 529 | *2.9* | ▲ |
| Denmark | 528 | *2.8* | ▲ |
| Germany | 528 | *2.9* | ▲ |
| Italy | 524 | *2.7* | ◆ |
| Portugal | 522 | *3.9* | ◆ |
| Slovenia | 520 | *2.7* | ◆ |
| N. Ireland | 517 | *2.6* | ◆ |
| **Ireland** | **516** | *3.4* | |
| Croatia | 516 | *2.1* | ◆ |
| Australia | 516 | *2.8* | ◆ |
| Serbia | 516 | *3.1* | ◆ |
| Lithuania | 515 | *2.4* | ◆ |
| Belgium (Fl) | 509 | *2.0* | ◆ |
| Romania | 505 | *5.9* | ◆ |
| Spain | 505 | *3.0* | ▼ |
| Poland | 505 | *2.6* | ▼ |
| **Centrepoint** | **500** | – | ▼ |
| New Zealand | 497 | *2.3* | ▼ |
| Kazakhstan | 495 | *5.1* | ▼ |
| Norway | 494 | *2.3* | ▼ |
| Chile | 480 | *2.4* | ▼ |
| Thailand | 472 | *5.6* | ▼ |
| Turkey | 463 | *4.5* | ▼ |
| Georgia | 455 | *3.8* | ▼ |
| Iran | 453 | *3.7* | ▼ |
| Bahrain | 449 | *3.5* | ▼ |
| Malta | 446 | *1.9* | ▼ |
| Azerbaijan | 438 | *5.6* | ▼ |
| Saudi Arabia | 429 | *5.4* | ▼ |
| UAE | 428 | *2.5* | ▼ |
| Armenia | 416 | *3.8* | ▼ |
| Qatar | 394 | *4.3* | ▼ |
| Oman | 377 | *4.3* | ▼ |
| Kuwait | 347 | *4.7* | ▼ |
| Tunisia | 346 | *5.3* | ▼ |
| Morocco | 264 | *4.5* | ▼ |
| Yemen | 209 | *7.3* | ▼ |

### Table Key

- Significantly above scale centrepoint
- At / near centrepoint
- Significantly below scale centrepoint
- ▲ Significantly higher than Ireland
- ◆ Not significantly different to Ireland
- ▼ Significantly lower than Ireland

**Table 2.2: Mean scores and standard errors for each domain in PIRLS and TIMSS 2011, and position relative to the study centrepoints and Irish means, out-of-grade tests and benchmarking participants**

| Reading | Mean | *SE* | IRL |
|---|---|---|---|
| **Sixth grade participants** | | | |
| Honduras | 450 | *4.8* | ▼ |
| Morocco | 424 | *3.9* | ▼ |
| Kuwait | 419 | *5.2* | ▼ |
| Botswana | 419 | *4.1* | ▼ |
| **Benchmarking participants** | | | |
| Florida, US | 569 | *2.9* | ▲ |
| Ontario, Canada | 552 | *2.6* | ◆ |
| Alberta, Canada | 548 | *2.9* | ◆ |
| Quebec, Canada | 538 | *2.1* | ▼ |
| Andalusia, Spain | 515 | *2.3* | ▼ |
| Dubai, UAE | 476 | *2.0* | ▼ |
| Maltese - Malta | 457 | *1.5* | ▼ |
| Abu Dhabi, UAE | 424 | *4.7* | ▼ |
| RSA (5th grade English/Af'aans) | 421 | *7.3* | ▼ |

| Maths | Mean | *SE* | IRL |
|---|---|---|---|
| **Sixth grade participants** | | | |
| Botswana | 419 | *3.7* | ▼ |
| Honduras | 396 | *5.5* | ▼ |
| Yemen | 348 | *5.7* | ▼ |
| **Benchmarking participants** | | | |
| N. Carolina, US | 554 | *4.2* | ▲ |
| Florida, US | 545 | *2.9* | ▲ |
| Quebec, Canada | 533 | *2.4* | ◆ |
| Ontario, Canada | 518 | *3.1* | ▼ |
| Alberta, Canada | 507 | *2.5* | ▼ |
| Dubai, UAE | 468 | *1.6* | ▼ |
| Abu Dhabi, UAE | 417 | *4.6* | ▼ |

| Science | Mean | *SE* | IRL |
|---|---|---|---|
| **Sixth grade participants** | | | |
| Honduras | 432 | *5.8* | ▼ |
| Botswana | 367 | *5.5* | ▼ |
| Yemen | 345 | *7.0* | ▼ |
| **Benchmarking participants** | | | |
| Florida, US | 545 | *3.7* | ▲ |
| Alberta, Canada | 541 | *2.4* | ▲ |
| N. Carolina, US | 538 | *4.6* | ▲ |
| Ontario, Canada | 528 | *3.0* | ▲ |
| Quebec, Canada | 516 | *2.7* | ◆ |
| Dubai, UAE | 461 | *2.3* | ▼ |
| Abu Dhabi, UAE | 411 | *4.9* | ▼ |

### Table Key

- Significantly above scale centrepoint
- At / near centrepoint
- Significantly below scale centrepoint
- ▲ Significantly higher than Ireland
- ◆ Not significantly different to Ireland
- ▼ Significantly lower than Ireland

With a mean score of 552, Irish pupils performed very well on the PIRLS assessment. Ireland's reading score is shaded green, meaning that it is significantly above the PIRLS scale centrepoint. Ireland is in tenth position in Table 2.1. However, only a few points separate Ireland from the four preceding countries – not a meaningful difference. Ireland's score was significantly lower than that of only five countries (the top four performers and Northern Ireland), indicated by ▲ in the IRL column. Ireland's mean score does not differ significantly from eight other countries (US, Denmark, Croatia, Chinese Taipei, England, Canada, Netherlands, and the Czech Republic). Thirty-one other countries – including Germany, Sweden, Australia and New Zealand – were significantly outperformed by Ireland, indicated by ▼ in the IRL column.

### Gender differences in reading

As in previous cycles of PIRLS, girls significantly outperformed boys. The international average was a 17-point gap in favour of girls (520 versus 504, after rounding). In six of the 45 countries (Colombia, Italy, France, Spain, Belgium [French-speaking area] and Israel) the gender difference was not significant. In *all* other participating countries, girls scored significantly higher than boys. Gender differences tended to be most pronounced in Arabic-speaking countries, including Saudi Arabia, Oman and Qatar.

In Ireland, girls averaged 559 on PIRLS, while boys averaged 544. The 15-point gap is statistically significant, and similar to the overall international average gender gap.

## Mathematics

The highest performing countries on the mathematics test were Singapore, the Republic of Korea and Hong Kong. Not only did these countries significantly outperform all of the 47 other participating countries, but they achieved national means in excess of 600 – more than a full standard deviation

above the scale centrepoint. With the exception of Chinese Taipei and Japan, all other countries lagged behind the three top performers by at least 40 points. Northern Ireland's mean of 562 meant that it was the highest performing country from outside the Asia-Pacific region.

Ireland's mean score of 527 is significantly above the TIMSS mathematics centrepoint of 500. Ireland is in 17th position in the TIMSS mathematics table. Irish pupils were significantly outperformed by pupils in 13 countries, including Northern Ireland, Finland, England and the US. Ireland's mean score for TIMSS mathematics does not differ significantly from the means for Lithuania, Portugal and Germany, and is significantly higher than the mean for 33 countries, including Australia, New Zealand, Italy and Sweden.

### Gender differences in mathematics

Previous cycles of TIMSS have found only small gender differences in overall mathematics performance at Fourth grade – a slightly larger gap, favouring girls, is apparent at Eighth grade – with no clear pattern across countries. This was also true in 2011. There was no significant gender difference for the overall international average, as girls averaged 490 and boys averaged 491. At country level, 26 countries had no significant gender differences, 20 had a small gap in favour of boys, and four (Qatar, Thailand, Oman and Kuwait) had a slightly larger gap in favour of girls. In Ireland, girls averaged 526 and boys averaged 529. The 3-point gap is not statistically significant, and is broadly similar to the overall international average gender gap.

## Science

On the TIMSS science assessment, the Republic of Korea and Singapore significantly outperformed all other countries, with means of 587 and 583, respectively. The next highest country was Finland with a mean of 570, significantly higher than all countries other than the top two. The means for all remaining countries were considerably lower – Japan, at 559, was closest to the top performers.

Ireland's overall score of 516 is significantly above the scale centrepoint of 500. Ireland is shown in 22nd position in the TIMSS science table. Taking measurement error into account, Ireland's mean score is significantly lower than that of 17 countries, including the US, Sweden, Netherlands, England, and Germany. Ireland's mean does not differ significantly from the means for 10 countries (including Italy, Northern Ireland and Australia), and is significantly higher than the mean for 22 countries, including Spain, New Zealand and Norway.

### Gender differences in science

As with mathematics, previous cycles of TIMSS have found only small gender differences in science at Fourth grade, with a slightly larger gap at Eighth grade. In 2011 there was no significant gender difference on overall international average science scores at Fourth grade – girls averaged 487 and boys averaged 485. At country level, 23 countries had no significant gender differences, 16 had a small gap in favour of boys, and three had a small gap in favour of girls. Eight countries – mainly Arabic-speaking countries – had larger gender differences favouring girls. In Ireland, boys and girls obtained identical mean scores (516).

# 3 Performance at International Benchmarks

As well as overall performance, both PIRLS and TIMSS report pupil achievement at four points on the scale, known as International Benchmarks. These enable descriptions of what pupils reaching each Benchmark can do. Benchmarks are cumulative skill sets, meaning that a pupil who is classified as at the Advanced Benchmark can also demonstrate all the skills that exemplify lower International Benchmarks. The descriptions of pupil skills were developed by international expert groups for each of the three domains, based on a detailed scale anchoring analysis. The cut-points, which are the same across all three tests and over different test cycles, are as follows:

| | |
|---|---|
| **625** | Advanced International Benchmark |
| **550** | High International Benchmark |
| **475** | Intermediate International Benchmark |
| **400** | Low International Benchmark |

This chapter describes, for each domain, the skills that pupils at each Benchmark can demonstrate. The skill descriptions will be enhanced if read in conjunction with the examples of actual test items and answers that are available on www.erc.ie/pirlstimss and http://timssandpirls.bc.edu/. The samples provide concrete examples of the types of questions that pupils at each International Benchmark can or cannot answer, based on performance in PIRLS and TIMSS 2011.

In addition, this chapter summarises the percentages of pupils in Ireland and in a selection of comparison countries who reached each of the International Benchmarks. The comparison countries comprised:

- the five other countries that conducted the assessment exclusively or almost exclusively in English (Australia, England, New Zealand, Northern Ireland, United States)
- two countries of interest due to very high all-round performance on PIRLS and TIMSS and/or PISA (Singapore and Finland)
- the three highest performing countries on a domain, if not already included for reasons outlined above.

Data are presented first for reading, then for mathematics, and finally, for science. Information on International Benchmark performance for other participating countries is available in the main international reports on the outcomes of PIRLS and TIMSS (Martin et al., 2012; Mullis et al., 2012a, 2012b).

## Reading Benchmarks

As noted earlier, reading texts were classified as either literary or informational, based on the reasons why a reader would engage with the text. Because of the very different demands each type of text presents, separate Benchmark descriptions are provided for literary and informational reading. The Benchmark descriptions outlined in Table 3.1 begin with the skills exemplifying the Advanced

International Benchmark. Pupils at the Advanced Benchmark can also display all the skills listed under lower Benchmarks. A progression in reading skills is evident across the Benchmarks. For example, a pupil at the Low International Benchmark is only able to retrieve specific and explicit elements of a text, whereas a pupil at the Advanced International Benchmark can use the text as a whole to support their interpretation of less tangible elements (such as a character's motivation).

**Table 3.1: Summary descriptions of the skills pupils display at each of the International Benchmarks for *reading*, by text purpose**

| Level | Literary Text | Informational Text |
|---|---|---|
| **Advanced (625)** | Integrate ideas and evidence across a text to appreciate overall themes.<br>Interpret story events and character actions to provide reasons, motivations, feelings, and character traits with full text-based support. | Distinguish and interpret complex information from different parts of text, and provide full text-based support.<br>Integrate information across a text to provide explanations, interpret significance, and sequence activities.<br>Evaluate visual and textual features to explain their function. |
| **High (550)** | Locate and distinguish significant actions and details embedded across the text.<br>Make inferences to explain relationships between intentions, actions, events, and feelings, and give text-based support.<br>Interpret and integrate story events and character actions and traits from different parts of the text.<br>Evaluate the significance of events and actions across the entire story.<br>Recognize the use of some language features (e.g., metaphor, tone, imagery). | Locate and distinguish relevant information within a dense text or a complex table.<br>Make inferences about logical connections to provide explanations and reasons.<br>Integrate textual and visual information to interpret the relationship between ideas.<br>Evaluate content and textual elements to make a generalization. |
| **Intermediate (475)** | Retrieve and reproduce explicitly stated actions, events, and feelings.<br>Make straightforward inferences about the attributes, feelings, and motivations of main characters.<br>Interpret obvious reasons and causes and give simple explanations.<br>Begin to recognize language features and style. | Locate and reproduce two or three pieces of information from within the text.<br>Use subheadings, text boxes, and illustrations to locate parts of the text. |
| **Low (400)** | Locate and retrieve an explicitly stated detail. | Locate and reproduce two or three pieces of information from within the text.<br>Use subheadings, text boxes, and illustrations to locate parts of the text. |

Content adapted from Exhibit 2.1, Mullis et al. (2012b).

A relatively high percentage of Irish pupils reached the Advanced International Benchmark (16%, compared to the international median of 8%) (Table 3.2). More than half of Irish pupils reached the High International Benchmark, and only 3% did not reach the Low International Benchmark (compared to the international median of 5%). Table 3.2 provides additional information about the distribution of achievement in Ireland and in the selected comparison countries.

Countries with higher mean scores tended to have larger percentages of pupils reaching the Advanced Benchmark. However, almost one quarter of Singaporean pupils reached the Advanced Benchmark, considerably higher than any other country. This is noteworthy because Hong Kong, the Russian Federation and Finland have (non-significantly) higher mean scores than Singapore, yet fewer students reaching the Advanced International Benchmark. On the other hand, 99% of pupils in Hong Kong, the Russian Federation and Finland reached the Low Benchmark, compared to 97% in Singapore.

**Table 3.2: National mean scores and percentages of pupils reaching the 2011 International Benchmarks for *reading*, Ireland and selected comparison countries**

| | National Mean | Percent of pupils (standard errors) | | | |
|---|---|---|---|---|---|
| | | Advanced | High | Intermediate | Low |
| Hong Kong | 571 | 18 (1.2) | 67 (1.5) | 93 (0.8) | 99 (0.2) |
| Russian Federation | 568 | 19 (1.2) | 63 (1.7) | 92 (1.1) | 99 (0.2) |
| Finland | 568 | 18 (0.9) | 63 (1.3) | 92 (0.7) | 99 (0.2) |
| Singapore | 567 | 24 (1.6) | 62 (1.8) | 87 (1.1) | 97 (0.4) |
| Northern Ireland | 558 | 19 (1.2) | 58 (1.4) | 87 (0.9) | 97 (0.6) |
| United States | 556 | 17 (0.7) | 56 (0.8) | 86 (0.6) | 98 (0.3) |
| Ireland | 552 | 16 (0.9) | 53 (1.4) | 85 (0.8) | 97 (0.5) |
| England | 552 | 18 (1.1) | 54 (1.3) | 83 (1.1) | 95 (0.5) |
| New Zealand | 531 | 14 (0.7) | 45 (1.1) | 75 (0.9) | 92 (0.5) |
| Australia | 527 | 10 (0.7) | 42 (1.1) | 76 (1.0) | 93 (0.7) |
| International median[4] | – | 8 (-) | 44 (-) | 80 (-) | 95 (-) |

## Mathematics Benchmarks

Table 3.3 provides summary descriptions of the skill-set that pupils at each International Benchmark for TIMSS mathematics can display. As was the case for reading, there is a clear progression of skills across the International Benchmarks. Pupils at the Advanced International Benchmark can *apply* knowledge in various complex settings and *explain* their reasoning. The complex skills they can demonstrate include solving multi-step word problems, applying geometric knowledge, and drawing conclusions from a table. At the other extreme, pupils at the Low International Benchmark can only demonstrate very basic mathematical knowledge. For example, they can add and subtract whole numbers, recognize some geometric shapes, and read simple tables.

The percentage of pupils in Ireland reaching the Advanced Benchmark in mathematics is more than twice the international median (9% versus 4%, respectively) (Table 3.4). However, it is well below the percentages (37-43%) in the top three performing countries of Singapore, the Republic of Korea, and Hong Kong, and also well below percentages for Northern Ireland (24%) and England (18%). On a more positive note, only 6% of pupils in Ireland fail to reach the Low International Benchmark (international median: 10%). The percentages reaching the Low and Intermediate Benchmarks are on a par with England, and are higher than the percentages in Australia and New Zealand. This suggests that Ireland's reasonably good performance on TIMSS mathematics can be attributed to having few very weak pupils, and a reasonable number of very advanced pupils.

[4] The values shown as the international median for each Benchmark are the percentages that divide countries evenly. For example, Advanced has a median value of 8%. This means that in half of participating countries more than 8% of pupils reached the Advanced Benchmark, and in the other half, fewer than 8% did so.

| Table 3.3: Summary descriptions of the skills pupils display at each of the International Benchmarks for *mathematics* | |
|---|---|
| | **Pupils at this Benchmark can...** |
| **Advanced (625)**<br>Apply mathematical knowledge in a variety of complex situations, and explain reasoning. | – solve a variety of multi-step word problems involving whole numbers, including proportions.<br>– show an increasing understanding of fractions and decimals.<br>– apply geometric knowledge of 2-D and 3-D shapes in various situations.<br>– draw a conclusion from data in a table, and justify the conclusion. |
| **High (550)**<br>Apply mathematical knowledge and understanding to solve problems. | – solve word problems involving operations with whole numbers<br>– use division in various problem situations, and use understanding of place value to solve problems.<br>– extend patterns to find a later specified term.<br>– demonstrate understanding of line symmetry and geometric properties.<br>– interpret and use data in tables and graphs to solve problems.<br>– use information in pictographs and charts to complete bar graphs. |
| **Intermediate (475)**<br>Apply basic mathematical knowledge in straightforward situations. | – show understanding of whole numbers, and some understanding of fractions.<br>– visualise 3-D shapes from 2-D representations.<br>– interpret bar graphs, pictographs, and tables to solve simple problems. |
| **Low (400)**<br>Have basic mathematical knowledge. | – add and subtract with whole numbers.<br>– have some recognition of parallel and perpendicular lines, familiar geometric shapes, and coordinate maps.<br>– read and complete simple tables and bar graphs |

Content adapted from Exhibit 2.1, Mullis et al. (2012a).

| Table 3.4: National mean scores and percentages of pupils reaching the 2011 International Benchmarks for *mathematics*, Ireland and selected comparison countries | | | | | |
|---|---|---|---|---|---|
| | **National Mean** | **Percent (standard errors)** | | | |
| | | **Advanced** | **High** | **Intermediate** | **Low** |
| Singapore | 606 | 43 (2.0) | 78 (1.4) | 94 (0.7) | 99 (0.2) |
| Republic of Korea | 605 | 39 (1.3) | 80 (0.8) | 97 (0.4) | 100 (0.1) |
| Hong Kong | 602 | 37 (1.8) | 80 (1.6) | 96 (1.0) | 99 (0.5) |
| Northern Ireland | 562 | 24 (1.3) | 59 (1.4) | 85 (1.2) | 96 (0.5) |
| Finland | 545 | 12 (0.8) | 49 (1.3) | 85 (1.2) | 98 (0.4) |
| England | 542 | 18 (1.3) | 49 (1.7) | 78 (1.4) | 93 (0.7) |
| United States | 541 | 13 (0.8) | 47 (1.1) | 81 (0.8) | 96 (0.3) |
| Ireland | 527 | 9 (0.9) | 41 (1.6) | 77 (1.4) | 94 (0.6) |
| Australia | 516 | 10 (0.9) | 35 (1.4) | 70 (1.4) | 90 (1.0) |
| New Zealand | 486 | 4 (0.5) | 23 (1.1) | 58 (1.3) | 85 (0.8) |
| International median | – | 4 (-) | 28 (-) | 69 (-) | 90 (-) |

## Science Benchmarks

Table 3.5 provides summary descriptions of the skill-set that pupils at each International Benchmark for TIMSS science can display. There are clear differences in the types of scientific knowledge and understanding displayed by pupils at each of the International Benchmarks. For example, pupils at the Advanced International Benchmark show some understanding of ecosystems, of organisms' interactions with the environment, and of the process of scientific inquiry. They are developing the skills to interpret results of a simple experiment, to draw conclusions from diagrams, and to evaluate and support an argument. In contrast, pupils at the Low International Benchmark show only basic

knowledge of life, physical, and earth sciences, and can only interpret simple diagrams or tables and provide short written responses to questions.

**Table 3.5: Summary descriptions of the skills pupils display at each of the International Benchmarks for *science***

| | Pupils at this Benchmark can... |
|---|---|
| **Advanced (625)**<br>Apply knowledge and understandings of scientific processes and relationships and show some knowledge of the process of scientific inquiry. | – communicate understanding of the characteristics and life processes of organisms, reproduction and development, ecosystems and organisms' interactions with the environment, and factors relating to human health.<br>– show an understanding of the properties of light and relationships among the physical properties of materials.<br>– apply and communicate their understanding of electricity and energy in practical contexts and demonstrate an understanding of magnetic and gravitational forces and motion.<br>– communicate their understanding of the solar system and of Earth's structure, physical characteristics, resources, processes, cycles, and history.<br>– have a beginning ability to interpret results in the context of a simple experiment, reason and draw conclusions from descriptions and diagrams, and evaluate and support an argument. |
| **High (550)**<br>Apply knowledge and understanding of the sciences to explain phenomena in everyday and abstract contexts. | – show some understanding of plant and animal structure, life processes, life cycles and reproduction, of ecosystems and organisms' interactions with their environment, including understanding of human responses to outside conditions and activities.<br>– demonstrate understanding of some properties of matter, electricity and energy, and magnetic and gravitational forces and motion.<br>– show some knowledge of the solar system, and of Earth's physical characteristics, processes, and resources.<br>– demonstrate elementary knowledge and skills related to scientific inquiry, compare, contrast, and make simple inferences, and provide brief descriptive responses combining knowledge of science concepts with information from both everyday and abstract contexts. |
| **Intermediate (475)**<br>Have basic knowledge and understanding of practical situations in the sciences. | – recognise some basic information related to characteristics of living things, their reproductive and life cycles, their interactions with the environment, and show some understanding of human biology and health.<br>– show some knowledge of properties of matter and light, electricity and energy, and forces and motion.<br>– know some basic facts about the solar system and show an initial understanding of Earth's physical characteristics and resources.<br>– demonstrate ability to interpret information in pictorial diagrams and apply factual knowledge to practical situations. |
| **Low (400)**<br>Show some elementary knowledge of life, physical, and earth sciences. | – demonstrate knowledge of some simple facts related to human health and the behavioural and physical characteristics of animals.<br>– demonstrate some basic knowledge of energy and the physical properties of matter.<br>– interpret simple diagrams, complete simple tables, and provide short written responses requiring factual information. |

Content adapted from Exhibit 2.1, Martin et al. (2012).

Table 3.6 shows the percentages of pupils in Ireland and in selected comparison countries who reached each of the four International Benchmarks for science. The proportion of pupils in Ireland who reached the Advanced Benchmark is slightly higher than the international median (7%, compared to 5% internationally). Slightly more pupils reached the Advanced and High Benchmarks in Ireland than in Northern Ireland or New Zealand, and the percentages in Ireland were similar to those in Australia. However, in the top three performing countries (Republic of Korea, Singapore,

and Finland), at least 20% of pupils reached the Advanced International Benchmark and at least 65% reached the High benchmark – markedly higher than the 35% of Irish pupils who reached the same benchmark. At 8%, the percentage of pupils in Ireland who failed to reach the Low International Benchmark was the same as the international median.

**Table 3.6: National mean scores and percentages of pupils reaching the 2011 International Benchmarks for *science*, Ireland and selected comparison countries**

| | National Mean | Percent (standard errors) | | | |
|---|---|---|---|---|---|
| | | Advanced | High | Intermediate | Low |
| Republic of Korea | 587 | 29 (1.5) | 73 (1.0) | 95 (0.4) | 99 (0.1) |
| Singapore | 583 | 33 (1.7) | 68 (1.7) | 89 (0.9) | 97 (0.4) |
| Finland | 570 | 20 (1.1) | 65 (1.7) | 92 (0.8) | 99 (0.3) |
| United States | 544 | 15 (0.8) | 49 (1.1) | 81 (0.8) | 96 (0.4) |
| England | 529 | 11 (0.9) | 42 (1.6) | 76 (1.3) | 93 (0.7) |
| Northern Ireland | 517 | 5 (0.6) | 33 (1.6) | 74 (1.3) | 94 (1.0) |
| Ireland | 516 | 7 (0.9) | 35 (1.7) | 72 (1.6) | 92 (0.9) |
| Australia | 516 | 7 (0.7) | 35 (1.4) | 72 (1.3) | 91 (1.0) |
| New Zealand | 497 | 5 (0.5) | 28 (1.1) | 63 (1.3) | 86 (0.9) |
| International median | – | 5 (-) | 32 (-) | 72 (-) | 92 (-) |

# 4 Spotlight on reading

As described in Chapter 1, there are two main elements to the PIRLS assessment: purpose and process. The two reading purposes – reading for literary experience or to acquire and use information – were each represented by five texts in the 2011 PIRLS assessment. The literary texts were intact short stories or episodes, averaging approximately 800 words, with supporting illustrations. Each text had two main characters and a plot. The informational texts contained a mixture of continuous and non-continuous texts, ranging between 600-900 words. These were accompanied by features such as diagrams, maps, or tables. Two purpose subscales were developed – Literary and Informational – corresponding to each of the two text types.

PIRLS is also based around four processes of reading comprehension, which were combined into two subscales for scaling purposes. Thus *retrieving explicitly stated information and making straightforward inferences* formed the Retrieve/Infer subscale, while *interpreting and integrating ideas and information* was combined with *examining and evaluating content, language, and textual elements* to form an Interpret/Evaluate subscale. Each pupil completed a booklet containing one Literary and one Informational text. The PIRLS 2011 assessment used 135 items, relatively evenly split between multiple-choice and constructed-response formats, and between the two process and two purpose subscales. The remainder of this chapter describes the performance of Ireland and comparison countries (see Chapter 3) – overall and by gender – on the reading subscales, and provides a broader context for the findings of PIRLS 2011.

## National differences in purpose and process scales

Generally, countries with the highest overall reading achievement tended to also have the highest achievement on the purpose and process subscales. However, many countries performed significantly higher or lower on one of the reading subscales than on the overall reading scale. Of the 45 participating countries, 30 obtained at least one purpose subscale score that differed significantly from their overall reading achievement score, while 29 differed significantly on at least one process subscale.

Table 4.1 presents information on subscale scores for Ireland and comparison countries. Only differences that are significantly different from the overall national means are shown. For example, in England there were no significant differences between the overall reading mean and the means for the purpose (Literary and Informational) subscales. However, the means on the process subscales did differ. England's mean for Retrieve/Infer was significantly poorer than their overall mean (by six points) while the mean for Interpret/Evaluate was significantly better than the overall mean (by four points).

Irish performance on the process subscales was broadly similar to performance on the overall scale. However, for the purpose subscales, Ireland's Literary score was 6 points higher than the overall reading mean, while the Informational subscale score was 3 points below the overall mean. Both differences are significant. With the exceptions of Finland and Australia, all comparison countries had some significant areas of relative strength or weakness on purpose or process, although the magnitude of the differences was typically relatively small. Among the comparison countries, the

largest difference observed was for the process subscales in Hong Kong, where pupils showed an almost 16-point advantage on Interpret/Evaluate over Retrieve/Infer.

**Table 4.1: National mean scores and statistically significant absolute differences from the mean for *reading purpose and process* subscales, Ireland and selected comparison countries**

| | National mean | Purpose | | Process | |
|---|---|---|---|---|---|
| | | Literary | Informational | Retrieve/ Infer | Interpret/ evaluate |
| Hong Kong | 571 | – 6 | + 7 | – 8 | + 7 |
| Russian Fed. | 568 | | | – 3 | + 2 |
| Finland | 568 | | | | |
| Singapore | 567 | | + 2 | | + 3 |
| Northern Ireland | 558 | + 5 | – 4 | – 3 | + 4 |
| United States | 556 | + 6 | – 4 | – 7 | + 6 |
| Ireland | 552 | + 6 | – 3 | | |
| England | 552 | | | – 6 | + 4 |
| New Zealand | 531 | + 2 | | – 4 | + 4 |
| Australia | 527 | | | | |

Only differences that are significantly different from the national mean are shown.

## Gender differences in purpose and process scales

Tables 4.2 and 4.3 present information on gender differences on subscale scores for Ireland and comparison countries. As there were significant gender differences on all subscales in each of the countries shown, and *all* in favour of girls, the convention of using bold font to denote significance is not used in these tables. It is worth noting, however, that gender differences were not significant in all participating countries. While girls significantly outperformed boys on the Literary subscale in all but two countries (Israel and Colombia), 11 countries (including Austria, France, Germany and Italy) did not have significant gender differences on the Informational subscale.

**Table 4.2: National overall mean reading scores, and gender gaps[5] (all in favour of girls) overall, and by *reading purpose* subscales, Ireland and selected comparison countries**

| | Overall | | Literary | | | Informational | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Gap | Girl | Boy | Gap | Girl | Boy | Gap |
| Hong Kong | 571 | 16 | 577 | 555 | 22 | 582 | 574 | 8 |
| Russian Fed. | 568 | 18 | 578 | 557 | 21 | 577 | 563 | 14 |
| Finland | 568 | 21 | 582 | 556 | 26 | 575 | 561 | 14 |
| Singapore | 567 | 17 | 578 | 556 | 22 | 576 | 563 | 13 |
| Northern Ireland | 558 | 16 | 575 | 552 | 23 | 561 | 549 | 12 |
| United States | 556 | 10 | 570 | 555 | 15 | 556 | 549 | 7 |
| Ireland | 552 | 15 | 569 | 546 | 23 | 553 | 545 | 8 |
| England | 552 | 23 | 567 | 539 | 28 | 560 | 539 | 21 |
| New Zealand | 531 | 20 | 546 | 521 | 25 | 537 | 522 | 15 |
| Australia | 527 | 17 | 539 | 516 | 23 | 534 | 522 | 12 |
| International Ave. | – | 17 | 522 | 502 | 20 | 519 | 507 | 12 |

All gender differences shown are statistically significant.

The overall international average for 2011 is not reported in Mullis et al. (2012b).

[5] As the international PIRLS database was not available at the time of writing, data are derived from rounded mean scores shown in Exhibits 1.5 and 3.7 in the main PIRLS report (Mullis et al., 2012b). This may lead to marginal rounding error in the reporting of the magnitude of the gender gap.

On average, the advantage for girls was more marked for literary than for informational texts (Table 4.2). For example, the international average gender difference on the Literary subscale was 20 points, compared to 12 points on the Informational subscale. In Ireland, there was a sizeable gender gap (23 points) on the Literary subscale, but only an 8-point gap for the Informational subscale. Irish boys obtained very similar scores on the Literary and Informational subscales (546 and 545, respectively). In contrast, for Irish girls, their performance on the Literary subscale was 16 points higher than on the Informational subscale (569 and 553, respectively). Gender differences were most apparent in England, as boys lagged behind their female counterparts by between 21 and 28 points, depending on the subscale.

The international average gender gaps for the two process subscales (Table 4.3) were quite similar (16 for Retrieve/Infer and 17 for Interpret/Evaluate). In Ireland, the equivalent gaps were 12 and 17 points, indicating gender differences on the reading process subscales that are broadly comparable to the international average.

**Table 4.3: National overall mean reading scores, and gender gaps (all in favour of girls) overall, and by *reading process* subscales, Ireland and selected comparison countries**

| | Overall | | Retrieve/Infer | | | Interpret/evaluate | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Gap | Girl | Boy | Gap | Girl | Boy | Gap |
| Hong Kong | 571 | 16 | 569 | 556 | 13 | 588 | 570 | 18 |
| Russian Fed. | 568 | 18 | 574 | 557 | 17 | 581 | 561 | 20 |
| Finland | 568 | 21 | 579 | 560 | 19 | 578 | 557 | 21 |
| Singapore | 567 | 17 | 573 | 557 | 16 | 579 | 562 | 17 |
| Northern Ireland | 558 | 16 | 563 | 548 | 15 | 571 | 553 | 18 |
| United States | 556 | 10 | 554 | 544 | 10 | 568 | 557 | 11 |
| Ireland | 552 | 15 | 558 | 546 | 12 | 562 | 545 | 17 |
| England | 552 | 23 | 557 | 535 | 22 | 568 | 544 | 24 |
| New Zealand | 531 | 20 | 536 | 519 | 17 | 545 | 526 | 19 |
| Australia | 527 | 17 | 536 | 517 | 19 | 538 | 521 | 17 |
| International Ave. | – | 17 | 521 | 505 | 16 | 519 | 502 | 17 |

All gender differences shown are statistically significant.
The overall international average for 2011 is not reported in Mullis et al. (2012b).

## Reading achievement in earlier studies

As noted earlier, 2011 was Ireland's first time to take part in PIRLS. As such, accurate trend data for reading are unavailable. However, Ireland did take part in the IEA's 1991 Reading Literacy Study. This, in conjunction with the results of successive National Assessments, can be used to make some broad comments on trends in overall reading performance and on areas of strength and weakness, subject to the following caveats. First, although there is a reasonable degree of overlap in the assessment frameworks used for the most recent National Assessments (2009) and PIRLS, the target grades are different (Fourth class for PIRLS, and Second and Sixth class for the National Assessments). Second, while the Reading Literacy Study provides data on reading achievement in Ireland relative to other countries, it used a slightly different assessment framework to that of PIRLS, and used two age-based cohorts (9- and 14-year-olds).

In addition to the Reading Literacy Study and the National Assessments, Ireland's participation in successive cycles of the OECD's Programme for International Student Assessment (PISA) allows general comparisons to be made with the performance of Irish post-primary students. The remainder

of this section is devoted to discussing Ireland's performance on each domain in terms of 1) overall national performance relative to the performance of other countries, and 2) strengths and weaknesses within Ireland on each domain (including gender differences).

### Overall national performance

Performance on PIRLS 2011 can be compared to the reading performance of 9-year-olds (all in Third class) in the IEA's Reading Literacy Study in 1991. Then, as now, the mean score of Irish pupils was significantly above the international average (Martin & Morgan, 1994). While it is inappropriate to compare Irish mean scores for reading across two different types of assessment – other than to note that in both instances, Ireland was above the overall study mean – we can examine Ireland's relative rank.

In 1991, 9-year-olds in Ireland obtained a mean reading score that was significantly *lower* than that of eight of the 27 countries that participated, and significantly *higher* than that of seven countries. In PIRLS 2011, Irish pupils obtained a mean reading score that was significantly *lower* than that of only five of the 45 countries that participated, and significantly *higher* than that of 31 countries. Thus, it seems reasonable to conclude that reading achievement in Irish primary schools has not deteriorated in the 1991-2011 period. However, it is not possible to draw firm conclusions about improved Irish performance due to the very different nature of the two studies and two sets of participating countries.

At post-primary level, Irish 15-year-olds performed above the OECD average for reading achievement on the first three cycles of PISA (Cosgrove, Shiel, Sofroniou, Zastrutzki & Shortt, 2005; Eivers, Shiel & Cunningham, 2008; Shiel, Cosgrove, Sofroniou & Kelly, 2001) but not on the most recent cycle (Perkins, Moran, Cosgrove & Shiel, 2010), where Irish reading performance did not differ significantly from the OECD average.

### Strengths and weaknesses

Reading was the only one of the three domains where a significant gender difference (in favour of girls) was found in Ireland. Comparing the PIRLS gender gap with the findings of other recent studies provides a somewhat mixed picture. In the 1991 Reading Literacy Study, Ireland had one of the largest gender gaps – again, favouring girls – among participating countries (Martin & Morgan, 1994). In contrast, the Irish gender gap in PIRLS 2011 is quite similar to the study average. All cycles of PISA have reported significantly higher reading achievement for Irish 15-year-old girls than for their male counterparts, with the size of the gap in Ireland broadly in line with the OECD average (Cosgrove et al., 2005; Eivers et al., 2008; Perkins et al., 2010; Shiel et al., 2001).

In contrast, the most recent National Assessments (Eivers et al., 2010) found mixed results with regard to gender differences on reading achievement (an advantage for girls at Second class, but a non-significant gap at Sixth class). Similarly, earlier cycles of the National Assessments and a recent replication of the National Assessments in Irish-medium schools also reported mixed results (Cosgrove, Kellaghan, Forde & Morgan, 2000; Eivers, Shiel, Perkins & Cosgrove 2005; Gilleece, Shiel, Clerkin & Millar, 2012). Girls generally obtained higher mean scores than boys on the assessments, but the size of the gap was often too small to be statistically significant.

In PIRLS 2011, Irish pupils displayed a relative strength on the Literary purpose subscale and a relative weakness on the Informational subscale. This seems primarily to be due to Irish girls' very good performance on Literary texts (a mean score of 569). In contrast, Irish boys performed at a

similar level on both types of text. The 1991 Reading Literacy Study provided somewhat similar findings (reporting on three text types: Narrative, Expository, and Documents) (Martin & Morgan, 1994). Irish 9-year-olds did best on Narrative texts, which is closest in type to the Literary subscale in PIRLS. Irish pupils showed a much poorer performance on Documents. Documents was composed of non-continuous texts only, and could be described as a subset of Informational texts. Performance by text type and gender is not available for Ireland, but, across the 9-year-old sample, gender differences were largest for Narrative and smallest for Document texts. In a related vein, PISA divides texts into continuous and non-continuous, with the latter representing the closest match to Informational texts. Performance on these reading subscales was reported for the cycles for which reading was the major PISA domain (2000 and 2009). While Irish 15-year-olds did not show major differences in performance across the two text types, the gender gap in favour of girls was larger for continuous texts (mainly, literary-type texts) than for non-continuous texts (Perkins et al., 2010; Shiel et al., 2001).

Irish pupils showed a relatively balanced performance by comprehension process (Retrieve/Infer versus Interpret/Evaluate), with no significant strengths or weaknesses on the process subscales. As the Reading Literacy Study did not provide subscales for processes, comparisons cannot be drawn. However, PISA reported on three reading process subscales (Retrieve, Interpret, and Reflect and Evaluate) in both 2000 and 2009. With the exception of a significantly higher mean score on Reflect and Evaluate in the 2009 cycle, Irish students tended to display reasonably similar levels of performance across process subscales.

# 5 Spotlight on mathematics

As described in Chapter 1, the TIMSS mathematics framework is organised around two dimensions: content and cognition. Within content, there are three domains, or areas of subject matter: Number, Geometric Shapes and Measures, and Data Display. There are also three cognitive domains – Knowing, Applying, and Reasoning – reflecting the range of thinking processes that pupils are expected to engage in as they answer the items presented in the assessment. Every item in the assessment was categorised by both cognitive and content domain. Half of the items were categorised as Number, 35% as Geometric Shapes and Measures, and 15% as Data Display items. In terms of cognition, 40% of items assessed Knowing, 40% assessed Applying, and 20% assessed Reasoning skills.

The mathematics component of the TIMSS 2011 assessment used 175 items, relatively evenly split between multiple-choice and constructed-response formats. Items were grouped into 14 blocks of items, and each pupil was presented with two of the 14 blocks, equivalent to half of each TIMSS test booklet. This chapter describes performance on the mathematics cognitive and content domains by pupils in Ireland and in a group of comparison countries (see Chapter 3 for more detail on comparison countries). Gender differences in domain performance are also discussed. Finally, Irish performance in TIMSS 2011 is placed in a wider context with comparison to previous national and international assessments of mathematics.

## National differences in the cognitive domains

Most countries showed significant variation in performance across the cognitive domains, suggesting areas of relative strength and weakness in their mathematics performance. Only eight of 50 countries displayed no significant differences across the three domains. Table 5.1 shows, for Ireland and selected comparison countries, the differences between the overall national mathematics score and each of the cognitive domain subscale scores. Only scores that are significantly above or below the overall national means are included. For example, relative to the Republic of Korea's national mean of 605, pupils achieved a significantly higher score (+9) on the Knowing subscale, a significantly lower score (–5) on the Applying subscale, and a Reasoning score that did not differ significantly from the overall national mean.

Irish pupils demonstrated a significant strength on the Knowing subscale, achieving a mean score (539) that was 12 points above the overall Irish mean for mathematics. Six of the nine comparison countries also showed a relative strength on the Knowing subscale, with the largest difference (23 points higher than the overall national mean) found in Singapore. Ireland's mean score of 529 on the Applying subscale was not significantly different to the national mean. In contrast, the Reasoning mean of 510 is 18 points lower than the overall Irish mean for mathematics and represents a relative weakness for Irish pupils.

Relative weaknesses in Reasoning were also evident in five of the nine selected comparison countries, namely, Singapore, Hong Kong, Northern Ireland, England, and the United States. As in Ireland, each of these countries had a relatively strong performance on Knowing compared to their overall national averages. In contrast, pupils in New Zealand achieved a lower score on Knowing and a

higher score on the Reasoning subscale. Six countries showed a significantly stronger or weaker performance on the third cognitive domain, Applying, but all by quite small margins.

**Table 5.1: National mean scores and statistically significant absolute differences from the mean for *mathematics cognitive domains*, for Ireland and selected comparison countries**

| | Overall | Knowing | Applying | Reasoning |
|---|---|---|---|---|
| Singapore | 606 | + 23 | – 4 | – 18 |
| Republic of Korea | 605 | + 9 | – 5 | |
| Hong Kong | 602 | + 17 | -4 | – 13 |
| Northern Ireland | 562 | + 17 | | – 25 |
| Finland | 545 | | | |
| England | 542 | + 10 | | – 11 |
| United States | 541 | + 15 | -2 | – 15 |
| Ireland | 527 | + 12 | | – 18 |
| Australia | 516 | | + 3 | |
| New Zealand | 486 | – 10 | + 4 | + 4 |

Only differences that are significantly different from the national mean are shown.

## Gender differences in the cognitive domains

Within-country gender differences for the three cognitive domains are shown in Table 5.2, with statistically significant differences indicated by the use of bold font. For example, in the Republic of Korea, boys significantly outperformed girls on the overall scale (by a seven-point margin), and on the Applying and Reasoning, but not on Knowing, subscales. *All* of the significant differences shown in Table 5.2 are in favour of boys. However, although not shown here, girls outperformed boys on at least one domain in seven of the 50 participating countries.

**Table 5.2: National overall mean mathematics scores and gender differences, and *mathematics cognitive domains* by gender, Ireland and selected comparison countries**

| | Overall | | Knowing | | Applying | | Reasoning | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Gap (B–G) | Girls | Boys | Girls | Boys | Girls | Boys |
| Singapore | 606 | – 4 | 631 | 627 | 603 | 600 | 591 | 585 |
| Rep. of Korea | 605 | **+ 7** | 613 | 616 | 597 | **602** | 597 | **608** |
| Hong Kong | 602 | **+ 6** | 618 | 620 | 594 | **600** | 584 | **593** |
| Northern Ireland | 562 | 0 | 578 | 582 | 566 | 564 | 538 | 537 |
| Finland | 545 | **+ 7** | 543 | **553** | 540 | **548** | 543 | 548 |
| England | 542 | + 3 | 550 | 554 | 540 | 544 | 529 | 533 |
| United States | 541 | **+ 9** | 550 | **561** | 534 | **543** | 523 | **528** |
| Ireland | 527 | + 3 | 539 | 540 | 528 | 530 | 507 | 512 |
| Australia | 516 | + 6 | 513 | 520 | 517 | 521 | 509 | **518** |
| New Zealand | 486 | 0 | 475 | 477 | 491 | 489 | 489 | 491 |
| International Ave. | – | + 1 | 492 | 492 | 488 | 489 | 487 | **489** |

Significant differences between genders, within country and domain, are shown in bold.
The overall international average for 2011 is not reported in Mullis et al. (2012a).

Although there was considerable variation in gender differences across countries, overall, boys and girls performed at a broadly similar level on each of the three cognitive domains. The gender gap for these subscales was much smaller than for the reading subscales, with the largest difference (two points) found on the Reasoning subscale. In Ireland, as well as in Singapore, Northern Ireland, England, and New Zealand, there were no significant gender differences on any of the three cognitive

domains. In contrast, boys significantly outperformed girls on the Applying subscale in the Republic of Korea, Hong Kong, Finland and the United States. A significant male advantage was found in the Republic of Korea, Hong Kong, and the United States on the Reasoning subscale, and in Finland and the United States for the Knowing subscale.

## National differences in the content domains

Only three of 50 participating countries – the Republic of Korea, Finland, and Croatia – displayed no statistically significant strengths or weaknesses on the mathematics content domains. Table 5.3 shows, for Ireland and selected comparison countries, the differences between the overall national mathematics score and each of the content domain subscales. Only scores that are significantly above or below the overall national means are included. Thus, the rows for the Republic of Korea and Finland are empty, as pupils in both countries demonstrated consistent levels of performance across all three content domains. In contrast, Singapore showed very large variation in performance across the different mathematics content areas. For example, there was a 31-point gap between the mean scores obtained by Singaporean pupils on the Number and Data Display subscales (619 and 588, respectively).

Irish pupils displayed a significant strength on the Number subscale, and relative weaknesses on Data Display and on Geometric Shapes and Measures. Pupils in Hong Kong and Northern Ireland were also relatively stronger on Number and weaker on Data Display, but did not share Irish pupils' relative weakness on Geometric Shapes and Measures. Australian pupils performed particularly well on Geometric Shapes and Measures, relative to their overall mathematics score. As a consequence of this, Australia obtained a mean score that was 14 points above Ireland's mean on this subscale (534 versus 520, respectively) despite Ireland having an overall mean score that was 11 points higher than that of Australia.

**Table 5.3: National mean scores and statistically significant absolute differences from the mean for *mathematics content domains*, Ireland and selected comparison countries**

| | Overall | Number | Geometric Shapes & Measures | Data Display |
|---|---|---|---|---|
| Singapore | 606 | + 13 | – 17 | – 18 |
| Republic of Korea | 605 | | | |
| Hong Kong | 602 | + 3 | + 3 | – 8 |
| Northern Ireland | 562 | + 4 | | – 8 |
| Finland | 545 | | | |
| England | 542 | – 3 | | + 7 |
| United States | 541 | + 2 | – 6 | + 4 |
| Ireland | 527 | + 5 | – 7 | – 4 |
| Australia | 516 | – 8 | + 18 | |
| New Zealand | 486 | – 3 | – 3 | + 5 |

Only differences that are significantly different from the national mean are shown.

## Gender differences in the content domains

Ireland was one of 15 countries (including Singapore, Northern Ireland, England and Australia) where no significant gender differences were found between overall mathematics mean score and the mean score on any of the content domains. In countries where significant gender differences were found, there tended to be an advantage for boys on Number (averaging about three points) (Table 5.4). There

was an overall advantage for girls on the Data Display subscale (a difference of about four points) and on Geometric Shapes and Measures (two points). Among our comparison countries, the United States displayed the most pronounced gender differences – a nine-point overall difference in favour of boys, and a significant male advantage on each of the three content domains.

**Table 5.4: National overall mean mathematics scores and gender differences, and *mathematics content domains* by gender, Ireland and selected comparison countries**

| | Overall | | Number | | Geometric Shapes & Measures | | Data Display | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Gap (B–G) | Girls | Boys | Girls | Boys | Girls | Boys |
| Singapore | 606 | – 4 | 621 | 617 | 591 | 588 | 591 | 584 |
| Republic of Korea | 605 | **+ 7** | 600 | **610** | 606 | 608 | 607 | 599 |
| Hong Kong | 602 | **+ 6** | 600 | **608** | 600 | **609** | 593 | 593 |
| Northern Ireland | 562 | 0 | 566 | 567 | 561 | 559 | 558 | 552 |
| Finland | 545 | **+ 7** | 538 | **552** | 544 | 543 | 549 | 553 |
| England | 542 | + 3 | 536 | 542 | 544 | 547 | 551 | 547 |
| United States | 541 | **+ 9** | 538 | **548** | 531 | **539** | 542 | **547** |
| Ireland | 527 | + 3 | 530 | 535 | 519 | 521 | 524 | 522 |
| Australia | 516 | + 6 | 505 | 511 | 532 | 536 | 512 | 519 |
| New Zealand | 486 | 0 | 481 | 485 | 482 | 484 | **496** | 487 |
| International Ave. | – | + 1 | 493 | **496** | **485** | 483 | **486** | 482 |

Significant differences between genders, within country and domain, are shown in bold.
The overall international average is not reported in Mullis et al. (2012a).

## Mathematics achievement in earlier studies

As noted earlier, Ireland has not participated in a large international comparative assessment of mathematics achievement at primary level since 1995. Therefore, the only direct comparisons that can be drawn are to the Irish results in mathematics from TIMSS 1995. However, three cycles of TIMSS have been completed since then, without Ireland's participation. Analysis of changes in Irish performance between cycles may not, therefore, be as informative as would have been the case were links to more recent cycles of assessment also available. That broad caveat aside, the link between Irish performance on TIMSS in 1995 and 2011 is reliable enough to allow some trend comparisons to be made.

This section also draws on three other sources – the Second International Assessment of Educational Progress (IAEP II), the results of successive National Assessments, and successive cycles of PISA. Conducted in 1991, IAEP II assessed the mathematics and science achievements of 9- and 13-year-old children in 20 countries, including Ireland. The 9-year-old cohort is most relevant for comparison with TIMSS. Ireland's National Assessments provide more recent comparisons to mathematics performance at primary level, while PISA provides information on the performance of Irish 15-year-old students in an international context. However, due to key differences (e.g., in target grades or ages or in underlying assessment frameworks), these studies can be used for broad comparisons only.

The remainder of this section is devoted to discussing Ireland's performance on each domain in terms of 1) overall national performance relative to the performance of other countries, and 2) strengths and weaknesses within Ireland on each domain (including gender differences).

## Overall national performance

Comparing Ireland's performance over two cycles of TIMSS, we find that the mean mathematics score achieved by Irish pupils increased from 523 points in 1995 to 527 points in 2011 (Table 5.5). This 5-point difference (before rounding) is not statistically significant. The mean score obtained by Irish girls increased by 1 point over the 16 years (from 525 to 526) while boys' mean score increased by 8 points from 521 to 529. While the increase for boys was larger than for girls, neither was statistically significant. By way of comparison, 16 other countries have reliable data for TIMSS in both 1995 and 2011. Of these, 12 achieved a significantly higher score on the mathematics assessment in 2011, three achieved a significantly lower score in 2011, and one showed no overall change.

Although there was no improvement in Ireland's *overall* mean score for mathematics, there was improved performance among low-achieving pupils. While the proportion of Irish pupils reaching the Advanced and High Benchmarks is similar in both 1995 and 2011, significantly more Irish pupils reached the Low Benchmark in 2011 (Table 5.5). In 1995, 9% of Irish pupils failed to reach the lowest Benchmark, while in 2011, this was reduced to only 6% of pupils (compared to a study median of 10%).

**Table 5.5: Overall mean score, and percentage of Irish pupils reaching the *mathematics* International Benchmarks, 1995 and 2011**

| | Mean | Advanced (625) | High (550) | Intermediate (475) | Low (400) |
|---|---|---|---|---|---|
| Ireland: 1995 | 523 | 10% | 40% | 73% | 91% |
| Ireland: 2011 | 527 | 9% | 41% | 77% | **94%*** |

* Significantly higher in 2011 than in 1995.

Irish 9-year-olds assessed in 1991 as part of IAEP II performed just below the average for the 14 countries that participated at that age level (Martin, Hickey & Murchan, 1992). On average, Irish pupils responded correctly to 60% of items, compared to an IAEP average of 63%. In Ireland, as in most participating countries, no significant gender differences for mathematics were found. At post-primary level, the first three cycles of PISA found that Irish 15-year-olds performed at the OECD average for mathematics achievement, and that Irish boys significantly outperformed Irish girls (Cosgrove et al., 2005; Eivers et al., 2008; Shiel et al., 2001). However, on the most recent cycle (Perkins et al., 2010), Irish performance in 2009 was below the OECD average, and no significant gender difference was found on mathematics.

## Strengths and weaknesses

In terms of mathematics content areas, the strong performance of Irish pupils in TIMSS 2011 on the Number content domain and relatively poorer performance on Geometric Shapes and Measures is broadly in line with the findings of TIMSS 1995. The 1995 and 2011 assessments used slightly different frameworks, so direct comparison by content area is not possible. That aside, the Number subscale covers similar content to two subscales used in 1995 – Fractions and Proportionality, and Whole Numbers. The former was a relative national strength in 1995, while performance on the latter was similar to the overall Irish mean score. Similarly, Geometric Shapes and Measures corresponds to two subscales in 1995 (Geometry, and Measurement, Estimation and Number Sense), both of which were areas of relative weakness in 1995 (Mullis et al., 1997).

The National Assessments in Ireland report separately on Measures and on Shape and Space content domains – these correspond to the combined Geometric Shapes and Measures domain in TIMSS 2011. As in TIMSS, poor performance on items assessing Measures has been a feature of previous National Assessments (Shiel, Surgenor, Close, & Millar, 2006). However, significant improvements on Shape and Space were found between 1999 and 2004 (Shiel et al., 2006). For the 2009 National Assessments (Eivers et al., 2010) the target grade levels changed from Fourth class to Second and Sixth class, meaning that more recent trends are unavailable. Data from PISA 2003 – the only time that mathematics has been the major PISA domain – also suggest that Shape and Space is a relative national weakness (Cosgrove et al., 2005).

In TIMSS 2011, performance on the Data Display subscale was also below the Irish overall mathematics average. This is somewhat surprising, as it contrasts with the TIMSS 1995 finding that Data Representation, Analysis and Probability was a relative strength of Irish Fourth class children (Mullis et al., 1997). Irish pupils have also performed well on Data items in recent National Assessments (Eivers et al., 2010; Shiel et al., 2006). Further, in PISA 2003, Irish students scored *above* the OECD average on the Uncertainty subscale (which can be loosely compared to Data Display in the TIMSS framework) (Cosgrove et al., 2005).

In terms of cognitive domains, Irish pupils showed a relatively weak performance on the Reasoning subscale, the most cognitively demanding of the three TIMSS domains. Information on performance by cognitive domains were not reported for TIMSS 1995 or PISA 2003, but are reported on in National Assessments. As with the content domains, however, there is no direct correspondence between the domains. Reasoning in TIMSS is similar to two National Assessments scales (Reasoning, and Applying and Problem-Solving). Performance on the Applying and Problem-Solving scale was identified as a relative weakness in the 2004 National Assessments, and while performance on the Reasoning subscale significantly improved between 1999 and 2004, this was partially offset by a small drop on Applying and Problem-Solving skills (Shiel et al., 2006).

# 6 Spotlight on science

As is the case for mathematics, the assessment of science is organised around two dimensions: content and cognition. Content for science is divided into three domains, Life science, Physical science, and Earth science. There are also three cognitive domains – Knowing, Applying, and Reasoning – reflecting the thinking processes that pupils are expected to engage in as they complete the assessment. Every item in the science assessment was categorised by both cognitive and content domain. As with the mathematics assessment, 40% of items assessed Knowing, 40% assessed Applying, and 20% assessed Reasoning skills. Content was weighted in favour of Life science (45% of items) and Physical science (35% of items), while 20% of items assessed Earth science.

The science component of the TIMSS 2011 assessment consisted of 172 items, relatively evenly split between multiple-choice and constructed-response formats. The science items were grouped into 14 blocks of items, and each pupil was presented with two of the 14 blocks, equivalent to half of each TIMSS test booklet. This chapter describes how pupils in Ireland and in a group of comparison countries (see Chapter 3) performed on the science cognitive and content domains. Gender differences in domain performance are also discussed. Finally, the performance of Irish pupils in TIMSS 2011 is placed in context with reference to previous assessments of science.

## National differences in the cognitive domains

Almost all participating countries showed relative strengths and weaknesses on the three cognitive domains. Of 50 countries, 46 performed significantly better or worse than their overall national science achievement score on at least one domain. Table 6.1 shows the differences between the overall national science mean score and each of the cognitive domain subscales for Ireland and the selected comparison countries. Only differences that are significantly above or below the overall national means are included. Thus, all cells are empty for Australia and New Zealand, as science achievement was relatively consistent in each cognitive domain.

**Table 6.1: National mean scores and statistically significant absolute differences from the mean for *science cognitive domains*, for Ireland and selected comparison countries**

| | Overall | Knowing | Applying | Reasoning |
|---|---|---|---|---|
| Republic of Korea | 587 | – 17 | + 7 | + 18 |
| Singapore | 583 | – 13 | + 6 | + 13 |
| Finland | 570 | + 9 | | – 10 |
| United States | 544 | + 2 | | – 7 |
| England | 529 | | + 4 | |
| Northern Ireland | 517 | | + 5 | – 14 |
| Ireland | 516 | | | – 7 |
| Australia | 516 | | | |
| New Zealand | 497 | | | |

Only differences that are significantly different from the national mean are shown.

In contrast, pupils in the Republic of Korea and Singapore showed substantial differences in each of the three cognitive domains. In both countries, Reasoning was a relative strength and Knowing was a relative weakness. In Ireland, and most of the other comparison countries, differences across the three cognitive domains were less marked. Nonetheless, Irish pupils' performance on the Reasoning subscale

was significantly poorer than on the overall science scale (509, compared to 516). With a subscale mean score of 503, pupils in Northern Ireland also demonstrated a relative weakness on Reasoning.

## Gender differences in the cognitive domains

Within-country gender differences for the three cognitive domains are shown in Table 6.2, with statistically significant differences shown in bold. For example, in the Republic of Korea, boys performed significantly better than girls overall and on the Knowing subscale, but there were no significant gender differences on the Applying or Reasoning subscales. In many countries, gender differences on the three cognitive domains were small. Ireland was one of 15 countries (including England, Australia, and New Zealand) where there were no significant gender differences on performance on any of the three cognitive domains.

As can be seen in Table 6.2, when the international average scores were calculated, girls significantly outperformed boys on Reasoning, while there was no significant gender difference for mean scores on either the Knowing or Applying subscales. Among our comparison countries, boys performed better than girls on two domains in the United States and on one domain in the Republic of Korea, while Singaporean girls were stronger than boys on one domain, and weaker on another. Gender differences (favouring boys) were most pronounced in the United States. The 10-point gender gap on the overall science scale in the United States can be attributed to relatively strong performance by boys on the Knowing and Applying domains.

**Table 6.2: National overall mean science scores and gender differences, and *science cognitive domains* by gender, Ireland and selected comparison countries**

| | Overall | | Knowing | | Applying | | Reasoning | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Gap (B–G) | Girls | Boys | Girls | Boys | Girls | Boys |
| Republic of Korea | 587 | **+ 8** | 563 | **576** | 590 | 597 | 604 | 606 |
| Singapore | 583 | + 4 | 565 | **574** | 586 | 592 | **601** | 592 |
| Finland | 570 | 0 | 580 | 579 | 569 | 568 | 559 | 561 |
| United States | 544 | **+ 10** | 541 | **551** | 537 | **552** | 537 | 537 |
| England | 529 | – 1 | 527 | 530 | 533 | 532 | 533 | 521 |
| Northern Ireland | 517 | – 1 | 518 | 517 | 520 | 523 | 505 | 500 |
| Ireland | 516 | + 1 | 516 | 520 | 516 | 518 | 513 | 505 |
| Australia | 516 | 0 | 515 | 520 | 513 | 513 | 520 | 515 |
| New Zealand | 497 | + 1 | 494 | 498 | 497 | 498 | 501 | 492 |
| International Ave. | – | –2 | 486 | 485 | 485 | 484 | **485** | 478 |

Significant differences between genders, within country and domain, are shown in bold.
The overall international average is not reported in Martin et al. (2012).

## National differences in the content domains

As with the cognitive domains, national strengths and weaknesses were evident on the three science content domains of Life science, Physical science, and Earth science. Forty-six (of 50) countries achieved a content subscale score that was significantly better or worse than the overall national mean on at least one of the three domains. Table 6.3 shows where significant differences were found between the overall national mean and the content domain scores for Ireland and the comparison countries. In Ireland and Finland, performance on all three content areas was relatively consistent, with no domain scores differing significantly from the overall mean.

In contrast, differences were found for each of the three content domains in the Republic of Korea and in Singapore. However, while pupils in the Republic of Korea showed a relative weakness on Life science, pupils in Singapore performed substantially poorer (by 42 points) on Earth science items. Pupils in Northern Ireland, England, and the US were relatively weaker on Earth science, although English and American pupils also displayed relative strengths in one other content domain.

**Table 6.3: National mean scores and statistically significant absolute differences from the mean for *science content domains*, for Ireland and selected comparison countries**

| | Overall | Life science | Physical science | Earth science |
|---|---|---|---|---|
| Republic of Korea | 587 | – 16 | + 10 | + 16 |
| Singapore | 583 | + 14 | + 15 | – 42 |
| Finland | 570 | | | |
| United States | 544 | + 3 | | – 5 |
| England | 529 | | + 7 | – 7 |
| Northern Ireland | 517 | | | – 9 |
| Ireland | 516 | | | |
| Australia | 516 | | | + 4 |
| New Zealand | 497 | | – 3 | |

Only differences that are significantly different from the national mean are shown.

## Gender differences in the content domains

Overall, girls and boys performed at a similar level on each content domain in only eight of the 50 countries that participated in TIMSS at the Fourth grade. Across all countries, boys performed more strongly than girls on Physical science and Earth science items, while girls performed more strongly than boys on Life science items.

No significant gender differences on the science content domains were found in Ireland, or in England and Australia (Table 6.4). In contrast, boys achieved a higher subscale score than girls on each of the three domains in the United States, and on two of the three domains in the Republic of Korea. Finnish boys and girls outperformed each other on one domain apiece while, in Northern Ireland, girls performed better than boys on Life science, with little gender difference evident on the other two domains.

**Table 6.4: National overall mean science scores and gender differences, and *science content domains* by gender, for Ireland and selected comparison countries**

| | Overall | | Life science | | Physical science | | Earth science | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Gap (B–G) | Girls | Boys | Girls | Boys | Girls | Boys |
| Republic of Korea | 587 | **+ 8** | 570 | 572 | 591 | **602** | 596 | **610** |
| Singapore | 583 | + 4 | 598 | 597 | 596 | 601 | 536 | **546** |
| Finland | 570 | 0 | **580** | 569 | 564 | **572** | 562 | 569 |
| United States | 544 | **+ 10** | 544 | **550** | 538 | **550** | 531 | **547** |
| England | 529 | – 1 | 534 | 527 | 532 | 538 | 520 | 524 |
| Northern Ireland | 517 | – 1 | **523** | 514 | 519 | 522 | 503 | 512 |
| Ireland | 516 | + 1 | 514 | 511 | 516 | 518 | 518 | 522 |
| Australia | 516 | 0 | 518 | 513 | 512 | 516 | 516 | 523 |
| New Zealand | 497 | + 1 | 499 | 496 | 493 | 494 | 494 | **504** |
| International Ave. | – | –2 | **489** | 481 | 484 | **485** | 479 | **483** |

Significant differences between genders, within country and domain, are shown in bold.
The overall international average is not reported in Martin et al. (2012).

## Science achievement in earlier studies

As Ireland has not participated in a comparative international assessment of science achievement at primary level since TIMSS in 1995, the only direct comparisons that can be drawn are to the Irish results from that implementation of TIMSS. As with mathematics, analysis of trends in Irish performance may not be as informative as would have been the case were links to more recent cycles of assessment also available. Nonetheless, the link between Irish performance on TIMSS in 1995 and 2011 is reliable enough to allow trend comparisons to be made.

Unlike reading and mathematics, National Assessment data are unavailable for science. Therefore, this section only draws on two other data sources – IAEP II and successive cycles of PISA. In 1991, Ireland was one of 20 countries that participated in the IAEP II assessment of the mathematics and science achievements of 9- and 13-year-old children. The 9-year-old cohort is most relevant for comparison with TIMSS. PISA provides information on the performance of Irish 15-year-old students in an international context. However, due to differences in target grades and underlying assessment frameworks, both studies can be used for broad comparisons only.

### Overall national performance

For science, the difference between the mean score of 515 in 1995 and the 2011 score of 516 is negligible (Table 6.5). Further, Ireland's performance at the International Benchmarks was similar to performance in 1995, with no significant changes in the percentages of pupils reaching each Benchmark. Irish performance in science is therefore consistent from 1995 to 2011, both in terms of overall achievement scores and in the proportion of pupils attaining each International Benchmark. By comparison, of the 16 other countries that have reliable data for TIMSS in 1995 and 2011, eight achieved a significantly higher science score in 2011 and one achieved a significantly lower score in 2011. The remaining seven countries, like Ireland, showed no overall change in performance.

**Table 6.5: Overall mean score, and percentage of Irish pupils reaching the *science* International Benchmarks, 1995 and 2011**

| | Mean | Advanced (625) | High (550) | Intermediate (475) | Low (400) |
|---|---|---|---|---|---|
| Ireland: 1995 | 515 | 8% | 36% | 70% | 91% |
| Ireland: 2011 | 516 | 7% | 35% | 72% | 92% |

The Irish 9-year-olds who were assessed in 1991 as part of IAEP II performed significantly below the international average on science (Martin et al., 1992). They obtained a mean percent correct score of 57% for science, compared to a 14-country average of 62%. In eight of the 14 countries, including Ireland, boys significantly outperformed girls.

In contrast, Irish 15-year-olds have consistently performed slightly above the OECD average on all cycles of PISA, including 2006, when science was the major assessment domain (Eivers et al., 2008). Thus, while the slightly above average science performance of Irish pupils on TIMSS 1995 and 2011 contrasts with the below average performance on IAEP II, it is in agreement with the various cycles of PISA from 2000 to 2009.

### Strengths and weaknesses

Irish pupils in TIMSS 2011 performed at a similar level on each of the science content domains (Earth science, Life science, and Physical science). This differs somewhat from achievement in TIMSS

1995, when the performance of Fourth class pupils was weaker on Physical science items than on the other content domains (Martin et al., 1997). In 1995, Irish pupils also performed relatively well on a fourth content domain – Environmental Issues and the Nature of Science – that is not presented as a separate subscale in 2011.

IAEP II also indicated some Irish strengths and weaknesses, relative to the study averages. Ireland was below the international study average for three content domains (Life Science, Nature of Science, and Physical Science) but similar to the international average for Earth and Space Science (Martin et al., 1992). In contrast, in PISA 2006 Irish 15-year-old students achieved similar scores on each of three content domains in that assessment – Earth and Space Systems, Living Systems, and Physical Systems – which are analogous to the content domains in the TIMSS science framework (Eivers et al., 2008). This similarity in content domain performance is consistent with the pattern found among Irish Fourth class pupils in TIMSS 2011.

With regard to the science cognitive domains, the performance of Irish pupils on the higher-order Reasoning process of the TIMSS science assessment was relatively poor compared to achievement on the science Knowing and Applying subscales. As noted in the previous chapter, poor Irish performance on Reasoning skills was also a feature of the mathematics component of TIMSS 2011.

Separate results for these cognitive process scales were not reported for TIMSS in 1995 (Martin et al., 1997). Therefore, comparable information on achievement on science cognitive subscales is only available for PISA 2006, the only cycle thus far for which science was the major domain. In that study, Irish students performed significantly above the OECD mean score on two cognitive skill subscales: Identifying Scientific Issues (corresponding loosely to Knowing in the TIMSS assessment framework) and Using Scientific Evidence. Irish performance was lower on the third cognitive skill, Explaining Scientific Phenomena. The score of Irish students on the latter subscale did not differ significantly from the OECD average (Eivers et al., 2008), compared to the above-average performance on the other two cognitive subscales.

# 7 Summary

This chapter provides a brief summary of some of the key achievement-related findings from PIRLS and TIMSS 2011. It is divided into two sections, the first of which recaps Ireland's performance – overall and by gender – on each of the three domains (reading, mathematics and science). This section also refers to Irish performance in previous international studies. The second section outlines the contents of a series of forthcoming thematic national reports. These reports will provide a more complete analysis of PIRLS and TIMSS 2011 from an Irish perspective. The thematic reports will draw on a more extensive range of data (including descriptions of cross-national differences in factors such as the home environment and school characteristics) than the present report, which deals almost exclusively with achievement outcomes.

## Summary of Irish performance in PIRLS and TIMSS 2011

This section summarises Ireland's performance on reading, mathematics and science, overall and by gender. Performance is compared against the 45 countries that participated in PIRLS and the 50 that took part in TIMSS. Where performance on subscales is significantly different from the overall scale, this is also noted in the text.

### Reading

Fourth class pupils in Ireland were among the top performers in the PIRLS assessment of reading. Their mean score of 552 was significantly above the scale centrepoint of 500. Ireland's mean score was significantly higher than that of 31 participating countries, and similar to that of eight other participants (including England, Denmark, Chinese Taipei, and the US). Only five countries – Hong Kong, the Russian Federation, Finland, Singapore, and Northern Ireland – obtained a mean score on the reading assessment that was significantly higher than that obtained by Ireland. Irish girls significantly outperformed boys on the reading assessment (559 compared to 544). Across PIRLS as a whole, girls significantly outperformed boys in 39 of the 45 participating countries.

Irish pupils displayed a relative strength on Literary texts. This was primarily due to Irish girls performing very well on items related to Literary texts, whereas Irish boys showed a good, balanced performance across both Literary and Informational texts.

As this is the first time Ireland has participated in PIRLS, no data are available for trend analyses. However, a very broad comparison of performance on the last major international comparative study of achievement in which Ireland took part (the Reading Literacy Study in 1991) suggests that reading achievement among Irish primary pupils has not deteriorated in the interim. Irish pupils' performance on PIRLS 2011 is – broadly – in line with performance on the 1991 Reading Literacy Study, and with the first three cycles of PISA, but not with the results of PISA 2009. The superior performance by Irish girls on PIRLS replicates the outcomes of the 1991 Reading Literacy Study and all cycles of PISA. However, a gender gap has not been a consistent feature of National Assessments in Ireland. As in PIRLS, literary-type texts proved a relative national strength in the 1991 Reading Literacy Study, but not in PISA 2000 or 2009. However, as in PIRLS, PISA did indicate that the gender gap in Ireland was larger for literary-type texts than for informational/non-continuous texts.

## Mathematics

On mathematics, Irish pupils achieved a mean score of 527, significantly above the scale centrepoint of 500. Ireland's mean score was significantly higher than the mean score of 33 other TIMSS participating countries, and similar to that of three participants (Lithuania, Portugal and Germany). Thirteen countries achieved mean scores that were significantly higher than Ireland's (including Singapore, Northern Ireland, Finland, England and the US).

In Ireland, girls averaged 526 and boys averaged 529, a non-significant gap that is of similar size to the international average gender difference (also not statistically significant). Some countries had significant gender differences for mathematics, but the gender gap tended to be small relative to the gender gap for reading. On the various mathematical content areas, Irish pupils displayed a relative strength on the Number subscale, and relative weaknesses on Data Display and on Geometric Shapes and Measures. On cognitive domains, Irish pupils showed a significant relative weakness on Reasoning (18 points lower than the overall mathematics score) and a relative strength on Knowing.

Ireland has not participated in TIMSS since 1995. Thus, trend data are available, but are not as robust as would be the case had Ireland participated in any of the intervening cycles. That caveat aside, trend analyses of the 1995 and 2011 data suggest no significant change in overall mathematics achievement among Irish Fourth class pupils since 1995. There was, however, a significant increase in the percentage of Irish pupils reaching the Low International Benchmark (i.e., improved performance among low-achieving pupils). In addition, Ireland was significantly above the TIMSS mathematics study centrepoint in TIMSS 1995 and 2011, in contrast to the performance on IAEP II in 1991.

As in PISA 2009 (but not any earlier PISA cycle) and IAEP II, TIMSS 2011 found no significant gender differences on mathematics performance among Irish pupils. The relative strength on the Number content domain and relative weakness on Geometric Shapes and Measures are both broadly in line with the findings of TIMSS 1995 and PISA 2003. However, the relative weakness on Data Display found in TIMSS 2011 is not reflected in other studies.

## Science

With a score of 516, Irish pupils performed significantly above the scale centrepoint of 500 on the TIMSS science assessment. Ireland's mean score was significantly higher than that of 22 other countries, and similar to the scores achieved in 10 countries, including Northern Ireland and Australia. Seventeen countries – including Singapore, Finland, the US, England, and Germany – had significantly higher mean scores on the science assessment than Ireland.

In Ireland, boys and girls obtained identical mean scores (516), mirroring the lack of any notable gender differences on the overall international average science scores – girls averaged 487 and boys averaged 485. Roughly half of participating countries had no significant gender differences on the science assessment. In countries where differences were found, those where girls outperformed boys were broadly counterbalanced by those where boys outperformed girls.

Irish pupils performed at a similar level across the three science content domains of Life science, Earth science and Physical science. The mean of 520 for Earth science was the highest score, but was not significantly higher than the means for Life and Physical science. However, on the cognitive

subscales for science, Irish pupils showed a relative weakness on Reasoning, with a mean score of 509, compared to their score of 516 on the overall science scale.

As with mathematics, Ireland has not participated in a comparative international study of science achievement at primary level since TIMSS 1995. Comparing science achievement across that timeframe, the difference between the mean score of 515 in 1995 and the 2011 score of 516 is negligible. Achievement by Irish pupils in science now is therefore broadly similar to performance in 1995. However, as with mathematics, it is worth noting that Ireland is significantly above the TIMSS science study centrepoint in TIMSS 1995 and 2011, in contrast to the performance on IAEP II in 1991, but similar to performance on successive cycles of PISA at post-primary level. The lack of any significant gender differences in science achievement among Fourth class pupils in 2011 also mirrors the findings of TIMSS 1995 and successive cycles of PISA from 2000 to 2009.

## Forthcoming reports

This volume, and the accompanying technical report for Ireland (Eivers & Clerkin, 2012) are the first in a series of national publications that present the findings of PIRLS and TIMSS 2011. A set of thematic reports is expected to follow in Spring 2013.

The forthcoming thematic reports will use the data arising from Ireland's participation in PIRLS and TIMSS 2011 – the first such internationally-comparative data to be available for Ireland at primary level since 1995 – to take a closer look at several topics of interest. The reports will contain *questionnaire* as well as achievement data, with a focus on Ireland, but in the context of data from other PIRLS and TIMSS participants. Topics for the thematic reports include:

- The characteristics of pupils who speak English (or, in other countries, the language of the test) as an additional language.
- The characteristics, attitudes, and classroom teaching practices of the teachers of Fourth grade pupils.
- Pupil engagement, attitudes to school and to school subjects, and relationships with peers.
- Interaction between the school and the home and parental involvement with school (including parents' and teachers' perceptions).
- Structural characteristics of the Irish education system, including class and school size, the time spent on various subject areas, and available resources.
- A multilevel model of achievement. This will provide more information on the relationships between pupil achievement and selected background variables.
- A review of some of the reading passages and items used in PIRLS 2011, with specific reference to how Irish pupils performed on the items, and how content relates to the teaching of reading in Ireland.
- A review of some of the mathematics items used in TIMSS 2011, with specific reference to how Irish pupils performed on the items, and how content relates to the mathematics curriculum in Ireland.
- A review of some of the science items used in TIMSS 2011, with specific reference to how Irish pupils performed on the items, and how content relates to the science curriculum in Ireland.

# References


Cosgrove, J., Kellaghan, T., Forde, P., & Morgan, M. (2000). *The 1998 National Assessment of English Reading*. Dublin: Educational Research Centre.

Cosgrove, J., Shiel, G., Sofroniou, N., Zastrutzki, S., & Shortt, F. (2005). *Education for life: The achievements of 15-year-olds in Ireland in the second cycle of PISA*. Dublin: Educational Research Centre.

Eivers, E. & Clerkin, A. (2012). *PIRLS and TIMSS 2011: Technical information for Ireland*. Dublin: Educational Research Centre.

Eivers, E., Close, S., Shiel, G., Millar, D., Clerkin, A., Gilleece, L., & Kiniry, J. (2010). *The 2009 National Assessments of Mathematics and English Reading*. Dublin: Educational Research Centre.

Eivers, E., Shiel, G., & Cunningham, R. (2008). *Ready for tomorrow's world? The competencies of Ireland's 15-year-olds in PISA 2006*. Dublin: Educational Research Centre.

Eivers, E., Shiel, G., Perkins, R., & Cosgrove, J. (2005). *The 2004 National Assessment of English Reading*. Dublin: Educational Research Centre.

Gilleece, L., Shiel, G., Clerkin, A., & Millar, D. (2012). *The 2010 National Assessments of English Reading and Mathematics in Irish-medium primary schools. Main report*. Dublin: Educational Research Centre.

Martin, M., Hickey, B., & Murchan, D. (1992). The Second International Assessment of Educational Progress: Mathematics and science findings in Ireland. *Irish Journal of Education, 26*, 3-146.

Martin, M. & Morgan, M. (1994). Reading literacy in Irish schools: A comparative analysis. *Irish Journal of Education, 28*, 3-101.

Martin, M., Mullis, I., Beaton, A., Gonzalez, E., Smith T., & Kelly D. (1997). *Science achievement in the primary school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Martin, M., Mullis, I., Foy, P., & Stanco, G. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M., Mullis, I., & Kennedy, A. (Eds.) (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I., Martin, M., Beaton, A., Gonzalez, E., Kelly, D., & Smith, T. (1997). *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Mullis, I., Martin, M., Foy, P., & Arora, A. (2012a). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I., Martin, M., Foy, P., & Drucker, K. (2012b). *PIRLS 2011 international results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I., Martin, M., Kennedy, A., Trong, K., & Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I., Martin, M., Minnich, C., Drucker, K., & Ragan, M. (2012c). *PIRLS 2011 encyclopaedia: Education policy and curriculum in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I., Martin, M., Minnich, C., Stanco, G., Arora, A., Centurino, V., & Castle, C. (2012d). *TIMSS 2011 encyclopaedia: Education policy and curriculum in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I., Martin, M., Ruddock, G., O'Sullivan, C., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Perkins, R., Moran, G., Cosgrove, J., & Shiel, G. (2010). *PISA 2009: The performance and progress of 15-year-olds in Ireland. Summary report*. Dublin: Educational Research Centre.

Shiel, G., Cosgrove, J., Sofroniou, N., & Kelly, A. (2001). *Ready for life: The literacy achievements of Irish 15-year olds with comparative international data*. Dublin: Educational Research Centre.

Shiel, G., Surgenor, P., Close, S., & Millar, D. (2006). *The 2004 National Assessment of Mathematics Achievement*. Dublin: Educational Research Centre.


ISBN 978-0-900440-40-3