

EDUCATIONAL RESEARCH CENTRE

NAMER 2021

The National Assessments of Mathematics and English Reading 2021

PERFORMANCE REPORT



JOANNE KINIRY, ALICE DUGGAN, ANASTASIOS KARAKOLIDIS, RACHEL
CUNNINGHAM AND DAVID MILLAR

Kiniry, Joanne.

The National Assessments of Mathematics and English Reading 2021: performance report/ Joanne Kiniry, Alice Duggan, Anastasios Karakolidis, Rachel Cunningham and David Millar.

Dublin: Educational Research Centre.

ix, 92 p.: illustrations, tables, graphs; 30cm.

ISBN: 978-1-911678-11-3.

1. National Assessments of Mathematics and English Reading (NAMER)
2. Educational surveys – Ireland
3. Educational tests and measurements – Ireland
4. Mathematics – Study and teaching (Primary) – Ireland
5. Mathematical ability – Testing – Ireland
6. Reading (Primary) - Ireland
7. Reading (Primary) – Ability testing – Ireland
8. COVID-19 Pandemic, 2020- - Educational aspects - Ireland

2023

I Title. II Kiniry, Joanne. III Duggan, Alice. IV Karakolidis, Anastasios. V Cunningham, Rachel. VI Millar, David.

372.48209417-dc23

How to cite this report:

Kiniry, J., Duggan, A., Karakolidis, A., Cunningham, R. & Millar, D. (2023). *The National Assessments of Mathematics and English Reading 2021: Performance Report*. Dublin: Educational Research Centre.

EDUCATIONAL RESEARCH CENTRE

NAMER 2021

The National Assessments of Mathematics and English Reading 2021

PERFORMANCE REPORT



Table of Contents

| | |
|---|------|
| Preface | V |
| Abbreviations | VII |
| Statistical Terms | VIII |
| Key Findings and Conclusions | 1 |
| Key Findings | 1 |
| Performance on Reading | 1 |
| Performance on Mathematics | 2 |
| Gender differences in performance | 3 |
| Conclusions | 4 |
| Performance on English reading and Mathematics | 4 |
| Gender differences | 4 |
| Literacy and Numeracy Strategy | 5 |
| COVID-19 pandemic | 5 |
| CHAPTER 1: Introduction | 6 |
| Adaptations to NAMER 2021 due to COVID-19 pandemic | 7 |
| Overview of previous cycles of NAMER | 8 |
| International assessments | 9 |
| PIRLS 2016 and PIRLS 2021 | 9 |
| TIMSS 2015 and TIMSS 2019 | 10 |
| PISA 2015 and PISA 2018 | 10 |
| Interim Review of the National Literacy and Numeracy Strategy | 12 |
| CHAPTER 2: Assessment Frameworks | 13 |
| English reading assessment framework | 14 |
| Introduction | 14 |
| Defining reading literacy/comprehension | 14 |
| Incorporating revisions to the syllabus into the framework | 15 |
| Reading vocabulary and comprehension in the framework | 16 |
| Specific elements of the NAMER 2021 framework for English reading | 17 |
| Correlations between reading subscales | 19 |
| Assessment development | 20 |
| Summary | 22 |
| Mathematics assessment framework | 22 |
| Correlations between mathematics content subscales | 25 |

| | |
|---|----|
| CHAPTER 3: Methods, Procedures and Technical Notes | 26 |
| Pilot study | 27 |
| Main study sample design and administration | 27 |
| Sample design | 27 |
| Response rates | 28 |
| Administration | 29 |
| Sampling variance | 30 |
| Sampling weights | 30 |
| Scaling of test data | 31 |
| Data analysis | 34 |
| CHAPTER 4: Achievement Outcomes and Trends | 36 |
| Second class English reading | 37 |
| Overall reading performance and performance by component | 37 |
| Reading performance by process skill | 39 |
| Proficiency levels | 40 |
| Sixth class Mathematics | 42 |
| Overall mathematics performance and performance by content area | 42 |
| Mathematics performance by process skill | 44 |
| Proficiency Levels | 46 |
| Summary | 49 |
| English reading | 49 |
| Mathematics | 49 |
| CHAPTER 5: Performance and Gender | 50 |
| Second class English reading performance | 51 |
| Performance on overall reading scale, by component and process skill | 51 |
| Proficiency levels | 54 |
| Trends in Second class English reading performance | 54 |
| Sixth class mathematics performance | 56 |
| Performance on overall mathematics scale, by content area and process skill | 56 |
| Proficiency levels | 59 |
| Trends in Sixth class mathematics performance | 59 |
| Summary | 62 |
| English reading | 62 |
| Mathematics | 62 |

| | |
|--|----|
| CHAPTER 6: Summary and Conclusions | 63 |
| Sample and tests | 64 |
| English reading | 64 |
| Mathematics | 65 |
| Gender differences | 66 |
| Gender and English reading | 66 |
| Gender and Mathematics | 66 |
| Conclusions | 67 |
| Performance on English reading and Mathematics | 67 |
| Gender gap | 68 |
| COVID-19 pandemic | 69 |
| Literacy and Numeracy Strategy | 69 |
| References | 71 |
| Appendices | 74 |
| Appendix for Chapter 4 | 74 |
| Second class English reading | 74 |
| Sixth class Mathematics | 78 |
| Appendix for Chapter 5 | 82 |
| Second class English reading | 82 |
| Sixth class Mathematics | 87 |

Preface

The National Assessments of Mathematics and English Reading (NAMER) aims to provide a snapshot of the mathematical and reading skills of Second and Sixth class pupils in Ireland, to identify areas of strength and weakness, and to inform educational policy and practice. Originally planned for administration in Spring 2020, NAMER was overtaken by the COVID-19 pandemic. The lockdowns necessitated by the pandemic led to extended school closures and the postponement of the National Assessments by a year. By the time NAMER was administered in May 2021, schools had been shut for a total of 96 days in the preceding two school years.

The National Literacy and Numeracy Strategy, originally published in 2011, set out performance targets for 2020 which were surpassed in the 2014 National Assessment. New performance targets in literacy and numeracy were set for 2020 in 2017. Although NAMER 2021 took place during a period of considerable upheaval in schools and society as a whole, the study provides an opportunity to monitor performance with reference to these targets.

The National Assessment in 2021

NAMER 2014 was the first National Assessment since 1980 in which statistically significant increases in pupil performance were observed. Other large-scale assessments conducted around the time of NAMER 2014 (e.g., PIRLS 2016, PISA 2015, TIMSS 2015) showed similar statistically significant gains in reading and mathematics performance, while more recent iterations of these studies (e.g., PISA 2018, TIMSS 2019, PIRLS 2021) have shown relative stability in performance since 2014. NAMER 2021 provided an opportunity to observe if similar patterns of performance in reading and Mathematics were continuing.

As in NAMER 2014, two reports are to be released examining the data collected in NAMER 2021, further reports will look at performance and background factors in schools participating in the Delivering Equality of Opportunity In Schools (DEIS) programme (Nelis & Gilleece, 2023). The current report focuses on the performance of pupils on the assessments of English reading and Mathematics administered as part of the National Assessments. A context report, to be published later in 2023, will situate performance in the context of the school and classroom practices, as well as pupil interests, experiences and background characteristics. The context report will, for example, link performance to teaching practices, class size and time allocated to teaching English reading and Mathematics in classrooms. The context report will explore some of the impact of the restrictions placed on schools by the COVID-19 pandemic.

In 2021, for the first time, NAMER included an extended sample of urban DEIS schools. Data from NAMER 2021 were used to write the 2021 DEIS report, which focuses on the performance of schools with designated DEIS status that participated in NAMER 2021 (Nelis & Gilleece, 2023).

Overview of the Performance Report

The current report comprises six chapters. The first chapter describes the broad context in which NAMER 2021 occurred, including the changes made to the administration of the study due to the impact of the COVID-19 pandemic, and changes since previous rounds of the National Assessments. Chapter 1 also considers the outcomes of recent international assessments, noting the strengths and weaknesses of Irish pupils at different levels of performance. The chapter outlines the review of the National Literacy and Numeracy strategy which led to the setting of new performance targets.

Chapter 2 describes the reading and mathematics frameworks used to assess pupil performance in NAMER 2021. Particular attention is given to the aligning of the English reading framework with the new Primary Language Curriculum (PLC).

Chapter 3 provides an overview of the methods and procedures used in the NAMER 2021 pilot and main studies. The design of the NAMER sample is discussed together with information on the administration of the study.

Chapter 3 also includes technical notes on the use of sampling weights and variance, as well as scaling of test data and data analysis.

Chapter 4 presents the achievement outcomes and trends for pupils who sat the Second class English reading assessment and Sixth class mathematics assessment. Overall performance is examined along with performance on identified subscales.

Chapter 5 presents achievement outcomes and trends in relation to gender. Performance and trends are again examined in terms of overall performance and performance on identified subscales.

Finally, Chapter 6 summarises this report, and draws broad conclusions

Acknowledgements

The Educational Research Centre (ERC) is proud to present this important study and hopes that it will be widely used and valued by educators, parents/guardians, and policy makers. We would like to express our sincere gratitude to all of the pupils, teachers, principals and schools who participated in this assessment, either in the pilot or main study, and contributed to its success. The ERC also acknowledges the support of Noreen Fiorentini and her colleagues in the Inspectorate of the Department of Education in overseeing the implementation of the National Assessments in a sample of schools.

The ERC wishes to acknowledge the support of members of the National Advisory Committee. As authors, we wish to acknowledge the support of colleagues at the Educational Research Centre, who provided support to the NAMER team during a very difficult and protracted project. These include: Aidan Clerkin (Acting CEO), who provided feedback on earlier drafts of this report; Jude Cosgrove (CEO until 2022); Gerry Shiel, who acted as Research Fellow until 2019; Lorraine Gilleece, Sharon Nelis and Caitríona Fitzgerald from the DEIS team who worked on the questionnaire materials and provided help and support with the administration of the study in 2020/2021; Anne Comey, Patricia Gaffney, Lynn Jackson, Imelda Pluck and Cian Ó Raghallaigh, who provided administrative and logistical support; all the casual staff members who made the study possible over the course of two years; John Coyle and Mary Delaney for their support with the selection of samples and the creation of databases; Seán Close, Brenda Donohue, Lynsey O’Keeffe and Sarah McAteer, who assisted with item development and who contributed to the pilot study in 2018/2019; Helena Ní Rócháin, Conall Ó Duibhir and Caitlin Myron who provided support with the translation and proofing of Irish language materials.

Finally, we wish to acknowledge the statistical support provided by Fernando Cartwright and the assistance of Carmen Robertson at Polymetrika International, Inc.

Joanne Kiniry
Alice Duggan
Anastasios Karakolidis
Rachel Cunningham
David Millar

Abbreviations

| | |
|--------------|---|
| CI | Confidence interval |
| CPD | Continuing Professional Development |
| CTT | Classical Test Theory |
| DEIS | Delivering Equality of Opportunity in Schools |
| EAP | Expected A Priori |
| ERC | Educational Research Centre |
| ICC | Item characteristic curve |
| ICT | Information and Communications Technology |
| IEA | International Association for the Evaluation of Educational Achievement |
| IRT | Item Response Theory |
| MML | Marginal maximum likelihood |
| MD | Mean difference |
| NAEP | National Assessment of Educational Progress |
| NAER | National Assessments of English Reading |
| NAMER | National Assessments of Mathematics and English Reading |
| NCCA | National Council for Curriculum and Assessment |
| OECD | Organisation for Economic Co-operation and Development |
| PIRLS | Progress in International Reading Literacy Study |
| PD | Percentage difference |
| PDST | Professional Development Service for Teachers |
| PISA | Programme for International Student Assessment |
| PLC | Primary Language Curriculum |
| PPS | Probability proportional to size |
| PSMC | Primary School Mathematics Curriculum |
| SD | Standard deviation |
| SE | Standard error |
| SED | Standard error of the difference |
| SEC | State Examinations Commission |
| SS | Scale score |
| TIMSS | Trends in International Mathematics and Science Study |

Statistical Terms

Percent correct scores

Percent correct scores indicate the proportion of items that each pupil answered correctly. The mean percent correct scores achieved by pupils in NAMER are provided for informational purposes only. These are based on item pools that have been modified between the three cycles of the assessment. Comparisons between groups across National Assessments, and within NAMER 2021, should not be made using percent correct scores. Instead, scale scores should be used.

Scale scores

Scale scores take into account not only the number of items answered correctly by each pupil but also the unique characteristics of each test item, as well as other information (e.g., contextual data). In NAMER 2009, mean scores on all scales and subscales in English reading and Mathematics were set to 250 points, and standard deviations were set to 50 points. Scores achieved by pupils participating in NAMER 2014 and 2021 were projected onto the same scales and subscales as those used in NAMER 2009 using Item Response Theory (IRT) scaling.

Standard deviation

The standard deviation is a measure of the dispersion (spread) of a set of data from its mean score. The more spread apart the data, the greater the deviation. In a normal distribution, 68 % of the scores are within one standard deviation of the mean, 95 % within two standard deviations, and 99 % within three.

Standard error

Estimates (e.g., mean scores and percentages) presented in this report are based on the sample of pupils selected to take part in NAMER. However, it is unlikely that the 'true' figure (e.g., the overall English reading mean score of all pupils in Ireland) would be exactly the same as the estimate calculated from our sample. Some variation or 'error' around estimates is to be expected. Thus, each estimate has a standard error, which provides information on how accurately the estimate found in our sample reflects the 'true' figure in the population. The 'true' figure is likely to be found in an interval that is about two standard errors on either side of the obtained estimate 95 % of the time, with a similar sample and assessment design.

Proficiency levels

Proficiency levels describe the skills that pupils falling within certain score ranges can demonstrate. There are four proficiency levels, with Level 4 representing the most complex skills and Level 1 the most basic. There is also a 'Below Level 1' category for pupils who did not show the competencies required for the simplest assessment tasks. Proficiency levels are based on mastery of skills, meaning that pupils are consistently able to demonstrate the skills at their proficiency level and the levels below, but are not consistently able to demonstrate the skills exemplifying the levels above them.

In NAMER 2009, pupils were assigned to proficiency levels on the overall reading and mathematics scales in Second and Sixth classes, such that, for each domain, at both class levels, 10 % of pupils were assigned to Level 4 (the highest level), 25 % to Level 3, 30 % to Level 2, 25 % to Level 1, and 10 % to 'below Level 1'. The same score benchmarks used in 2009 within each domain and grade level to assign pupils to the different proficiency levels were used in NAMER 2014 and 2021.

Statistical significance

If the difference between two estimates (e.g., mean scores) or the correlation between two variables is statistically significant, it means that the magnitude of the statistic is unlikely to occur if the relationship were not real. A statistical test has been carried out to determine if the probability of each statistic being observed by chance is less than 5 %. The criterion, or alpha level (α), of .05 (5 %) implies that only observed statistics with less than a 1 in 20 chance of occurring are interpreted as being statistically significant. It should be noted that 'significant' in the statistical sense does not necessarily mean significant (meaningful) in the everyday sense (see the section on 'effect sizes', below).

Confidence intervals

Confidence intervals provide a range of values within which a statistic of interest is expected to fall. In this report, we use the 95 % confidence intervals meaning that we are 95 % certain that the true value of an estimate would fall within the estimated confidence intervals around the observed value. To compute the confidence intervals around an estimate, the following formula is used:

$$CI = x \pm SE * t$$

Where x is the observed value (e.g., mean score or percentage), SE is the standard error around this estimate and t is the critical value which is based on the survey design and the significance level.

Correlation coefficients

Correlation coefficients describe the strength of a relationship between two variables on an interval or ordinal scale, such as between Reading Vocabulary and Reading Comprehension. The value of a correlation (the r value) ranges from -1 to +1. A positive correlation indicates that, as one variable increases, the other does too, while a negative correlation indicates that as one increases the other decreases. A correlation coefficient close to 0 indicates little or no relationship, while the closer the value is to ± 1 , the stronger the relationship. The following rule of thumb can be used to interpret the strength of the correlation coefficients reported in this volume:

| | |
|---------------------------|-------------------|
| Weak | < .1 |
| Weak to moderate | .1 to .24 |
| Moderate | .25 to .39 |
| Moderate to strong | .4 to .55 |
| Strong | >.55 |

Effect sizes

An effect size is a standardised measure of the strength of a relationship between two variables. If both variables have interval or ordinal scales, then the effect size is the correlation coefficient. If one variable describes membership in a group and the other has an interval or ordinal scale, then the effect size is the difference between two means that is expressed in standard deviation units. Mean difference effect sizes in this report were computed using Cohen's *d* (Cohen, 1988).

Based on benchmarks suggested by Cohen (1988), for mean differences, an effect size of 0.2 can be interpreted as small, an effect size of .5 is medium, and an effect size of .8 is large. However, these benchmarks should not always be interpreted rigidly, because even small effect sizes can have large consequences in some contexts.

This report uses the What Works Clearinghouse (2014) criteria for interpreting effect sizes. Mean differences with effect sizes of 0.25 or higher can be considered substantively important, whether or not the underlying difference is statistically significant. Effect sizes greater than 0.50 are considered large.

Key Findings and Conclusions

The 2021 National Assessments of Mathematics and English Reading (NAMER) were administered to representative samples of over 10,000 pupils in Second and Sixth classes in 188 primary schools during the first two weeks of May 2021. The study had been scheduled to take place in spring 2020 but was postponed due to the COVID-19 pandemic. Given the pandemic context, it was decided to adapt the study to reduce the administrative burden on schools and the testing load on pupils. In NAMER 2021, pupils in Second class took the English reading test only, and pupils in Sixth class took the mathematics test only. This differed from NAMER 2009 and 2014, in which pupils at both grade levels took both tests. It was also decided not to administer the parent/guardian questionnaire in 2021.

The pupils that participated in the tests had experienced two months of remote teaching and learning in January and February 2021, due to the COVID-19 pandemic. Pupils returned to face-to-face teaching in school in March 2021. Class teachers acted as test administrators and school coordinators acted as security/quality monitors. Inspectors of the Department of Education visited a sample of schools and oversaw testing procedures.

The current report focuses on the performance of pupils in NAMER 2021 and relates their performance to that of pupils who participated in the National Assessments in 2014. A second report, due to be published in late 2023, will examine factors relating to the teaching and learning of Mathematics and English reading, using school-level, classroom-level, and pupil-level data, drawn from questionnaires administered in conjunction with NAMER 2021.

The tests used in NAMER 2021 were secure curriculum-based instruments. The 2009 assessment framework, which had also been used in NAMER 2014 was adapted and extended ahead of the 2021 National Assessments to account for the introduction of the new Primary Language Curriculum (PLC). Small numbers of items in each test were dropped and replaced with items that had been trialled in the pilot study in 2019. There were multiple test booklets in each domain, allowing for greater coverage of content and processes. Item Response Theory (IRT) scaling was used to link booklets, and to allow for cross-cycle comparisons of pupil performance.

Findings related to the performance of pupils attending DEIS schools are not presented in this report.

These findings are described in a separate report focusing on pupil performance and background factors in DEIS schools (Nelis & Gilleece, 2023).

Key Findings

The following are the key findings from the performance report:

Performance on Reading

- Overall performance on reading in Second class was broadly similar to 2014. There was a small drop in performance between 2014 and 2021, however, it is not statistically significant. A similar pattern was observed in pupils' performance on the Vocabulary and Comprehension subscales and the Retrieve, Infer and Interpret & Integrate process skill subscales; performance was lower in 2021 than in 2014, but the changes in performance are not statistically significant.

- There are no statistically significant differences in overall reading scores at key percentile points between 2014 and 2021. Neither are there any statistically significant differences at key percentile points on the Vocabulary, Comprehension or Infer subscales. However, some statistically significant differences between 2014 and 2021 are observed at key percentile points on the Retrieve and Interpret & Integrate subscales. Specifically, scores at the 75th percentile on the Retrieve subscale and the 75th and 90th percentiles on the Integrate & Interpret subscales in 2021 were statistically significantly lower than the corresponding scores in 2014.
- The percentages of pupils performing at the different proficiency levels did not change statistically significantly in 2021, compared with 2014. The performance targets for Second Class reading set out in the Interim Review of the National Strategy to Improve Literacy and Numeracy, based on the proportions of pupils performing at the highest and lowest proficiency levels, were not met. However, the proportions of pupils performing at these levels in 2021 are similar to the corresponding proportions in 2014, and the 2021 results fall within the original targets set in the 2011 Strategy.

Performance on Mathematics

- Sixth class pupils' overall performance in Mathematics was broadly similar to 2014. Overall mathematics performance in NAMER 2021 was lower than in 2014, but the difference is not statistically significant. Performance on all content area subscales (Number & Algebra, Shape & Space, Measures, Data) and process skill subscales (Understand & Recall, Implement, Reason, Integrate & Connect, Apply & Problem Solve) was lower in 2021 compared with 2014, but only one subscale (Shape & Space) showed a statistically significant reduction in performance across cycles. For all other subscales, no statistically significant changes are observed in 2021, compared with 2014.
- There were no statistically significant differences in scores on the overall mathematics scale at key percentile points between 2014 and 2021. However, there were some statistically significant differences at key percentiles between 2014 and 2021. Scores at the 50th and 75th percentiles on the Shape & Space subscale were statistically significantly lower than the corresponding scores in 2014. On the Integrate & Connect subscale, the mean score at the 75th percentile was statistically significantly lower in 2021 than the corresponding score in 2014, while scores at the 25th and 90th percentiles were statistically significantly higher in 2021 than the corresponding scores in 2014.
- The percentages of pupils performing at the different proficiency levels in 2021 were not statistically significantly different from the percentages observed in 2014. The performance targets for Sixth Class Mathematics set out in the Interim Review of the National Strategy to Improve Literacy and Numeracy (Department of Education & Skills, 2017), based on the proportions of pupils performing at the highest and lowest

proficiency levels, were not reached. However, no statistically significant changes in performance are observed between 2014 and 2021 at each proficiency level overall, and the 2021 results are within the bounds of the original targets set in 2011.

Gender differences in performance

- In NAMER 2021, for the first time, pupils had three options for indicating their gender on the Pupil Questionnaire – Girl, Boy and Other. Small numbers of pupils at both grade levels selected Other. For this reason, this category was not used in statistical analyses. Comparative analyses were conducted only between girls and boys.
- In Second class English reading in NAMER 2021, girls outperformed boys. This is consistent with the pattern observed in 2014. In 2021, girls had higher scores than boys overall and across both subcomponents, and these differences are statistically significant for the overall scale and the Comprehension subscale. Boys and girls performed similarly to each other on the Vocabulary subscale. Girls' mean scores were higher than boys on the process skill subscales (Retrieve, Infer, Interpret & Integrate) but the difference is statistically significant only for the Infer subscale.
- In 2021, the proportion of girls in Second class achieving at the lower proficiency levels in English reading is statistically significantly smaller than the percentage of boys achieving at these low levels. In other words, there are more lower-achieving boys than girls. A greater percentage of girls than boys achieved at the higher proficiency levels but this difference is not statistically significant. There are no statistically significant changes in the proportion of boys and girls performing at each proficiency level in Second class reading in 2021, compared with 2014.
- In Sixth class Mathematics in NAMER 2021, boys performed at a higher level than girls overall and on all subscales. The differences are statistically significant on the overall mathematics scale and on two subscales (Measures and Number & Algebra). Boys also statistically significantly outperformed girls on three process skill subscales (Understand & Recall, Reason and Apply & Problem Solve). Girls and boys performed similarly to each other on the Shape & Space, Data, Implement and Integrate & Connect subscales.
- In terms of proficiency levels in Sixth class Mathematics, the proportion of boys achieving at or above Level 3 is statistically significantly higher than the corresponding proportion of girls. Conversely, the proportion of girls achieving at the lowest proficiency levels is statistically significantly higher than the percentage of boys achieving at these lowest levels. The proportions of boys and girls performing at the highest proficiency levels in 2021 decreased slightly compared with performance in 2014, but these decreases are not statistically significant. The proportion of girls performing at the lowest proficiency levels also increased slightly in 2021, but the increase is not statistically significantly different from the proportion of girls performing at the lowest levels in 2014.

Conclusions

Performance on English reading and Mathematics

- Second class pupils' performance in English reading and Sixth class pupils' performance in Mathematics in NAMER 2021 were broadly consistent with their counterparts' performance in 2014. It is noteworthy that NAMER 2014 was the first National Assessment since 1980 in which there were statistically significant increases in pupil performance. Given that NAMER 2021 was administered in the context of the COVID-19 pandemic, the relative stability of pupil performance at both grade levels is to be welcomed.
- Other large-scale assessments that were conducted around the time of NAMER 2014 (e.g., PIRLS 2016, PISA 2015) showed similar statistically significant gains in reading performance, while more recent iterations of these studies (e.g., PISA 2018) have shown slight (though not statistically significant) declines, which suggest stability in reading performance since 2014. The results of PIRLS 2021 also point towards stability in reading performance among Fourth class pupils since the previous cycle in 2016.
- The large increase in mathematics performance observed in NAMER 2014, compared with 2009, seemed to be reflected in a similar trend of increased performance in Mathematics in TIMSS 2015, compared with 2011. TIMSS 2019 results showed relative consistency in performance, similar to the stability observed in mathematics performance in NAMER 2021. This suggests that the gains in performance in NAMER 2014 have been maintained. The next cycle of TIMSS, due to take place in 2023, will offer more information on the mathematics performance of primary school pupils (Fourth class) in Ireland.

Gender differences

- A statistically significant performance gap in favour of girls was observed in Second class English reading in NAMER 2021. This is in keeping with the results of NAMER 2014. Other large-scale assessments (e.g., PIRLS 2016, PIRLS 2021, PISA 2015, PISA 2018) have shown a similar pattern of reading performance, with girls outperforming boys.
- In Sixth class Mathematics in NAMER 2021, a statistically significant performance gap in favour of boys was observed on the overall mathematics scale. This contrasts with NAMER 2014, when there was no statistically significant gender gap in overall mathematics performance at Sixth class. Likewise in TIMSS 2015 and TIMSS 2019, the gender differences in mathematics performance among Fourth class pupils were not statistically significant. The next cycle of TIMSS, due to take place in 2023, will provide more information on gender differences in performance in Mathematics at primary level.

Literacy and Numeracy Strategy

- The revised performance targets for English reading and Mathematics set in the Interim Review of the National Literacy & Numeracy Strategy were not reached in NAMER 2021. However, the results of NAMER 2021 are similar to those in NAMER 2014 and therefore fall within the original performance targets set in the 2011 Strategy. It is reassuring that results for both English reading performance at Second class and mathematics performance at Sixth class are broadly consistent with results from NAMER 2014 despite the significant challenges for the education system during the COVID-19 pandemic.
- Given the extended timeframe since the last cycle of the National Assessments, and in particular due to the pandemic-related school shutdowns and periods of remote teaching and learning that occurred in Spring 2020 and 2021, it is difficult to draw strong conclusions about the impact of specific elements of the Literacy and Numeracy strategy implemented since 2014.

COVID-19 pandemic

- Due to the COVID-19 pandemic, NAMER was postponed from 2020 to 2021, and it was decided to reduce the scope of the study (only one test was administered at each grade level and the parent questionnaire was not administered). Despite the unprecedented challenges of the pandemic and the difficulties faced by schools in administering the study, the overall participation rate was very high (91.5 %). Inspectors from the Department of Education visited a sample of schools (20 %) in 2021, while in previous cycles all schools would have been visited. Reports from schools visited by Inspectors suggest that there was strong adherence to the standardised assessment procedures detailed in the administration manuals, and that the assessments were kept secure.
- Due to the extended time period since NAMER 2014 and the cross-sectional design of the study, it is not possible to draw conclusions from NAMER 2021 about the specific impact of the pandemic on pupils' learning and performance.

CHAPTER 1

Introduction

The National Assessments have been conducted periodically in Ireland since 1972. They are an important source of information on national standards of achievement. The outcomes offer insights into pupils' strengths and weaknesses in particular domains and skills, and provide rich contextual information about home and school factors associated with achievement. Results of National Assessments have been used to inform policies that aim to improve the effectiveness of the educational system - for example, the National Strategy for Literacy and Numeracy (Department of Education and Skills, 2011).

Early cycles covered various class levels and domains but, since 2009, the assessments have focused on monitoring national standards in both Mathematics and English reading at Second and Sixth classes (Educational Research Centre, 2008, 2009; Eivers et al., 2009), hence the acronym NAMER (National Assessments of Mathematics and English Reading). The study involves the administration of tests of Mathematics and English reading to pupils in selected classes, as well as the administration of questionnaires to pupils, their parents/guardians, teachers and school principals. This report focuses on pupil achievement; factors related to achievement and contextual information gathered from the questionnaires will be examined in a subsequent context report. Tracking trends in performance is an important function of NAMER; therefore, throughout this report, results from NAMER 2009 and 2014 are presented alongside the current results from 2021, and comparisons are made between 2014 and 2021.

Adaptations to NAMER 2021 due to COVID-19 pandemic

The 2021 assessments were originally scheduled to take place in spring 2020 following a pilot study in 2019; however, due to the COVID-19 pandemic, schools in Ireland closed on March 12th, 2020, and did not reopen until September 2020. The decision to postpone NAMER until 2021 was taken in May 2020. In light of the challenging circumstances caused by the pandemic, the scope of the assessments was reduced to minimise the administrative load on school staff and the testing burden on pupils, while still gathering sufficient data for the study. In previous cycles of NAMER, pupils at both grade levels completed two tests – one in English reading and one in Mathematics. However, in this cycle, schools had just re-opened in March 2021 following a two-month period of remote teaching and learning; therefore it was decided that pupils in Second class would take the English reading test only, and pupils in Sixth class would take the mathematics test only. It was also decided not to administer questionnaires to parents/guardians of participating pupils, as had been done in previous cycles. Table 1.1 lists the instruments used in NAMER; '✓' indicates that the instrument was administered.

Table 1.1. Instruments used in NAMER

| Instrument | NAMER 2009 | NAMER 2014 | NAMER 2021 |
|---|------------|------------|------------|
| 2nd Class English Reading test | ✓ | ✓ | ✓ |
| 2nd Class Mathematics test | ✓ | ✓ | - |
| 2nd Class Pupil Questionnaire | ✓ | ✓ | ✓ |
| 2nd Class Teacher Questionnaire | ✓ | ✓ | ✓ |
| 2nd Class Parent/Guardian Questionnaire | ✓ | ✓ | - |
| 6th Class English Reading test | ✓ | ✓ | - |
| 6th Class Mathematics test | ✓ | ✓ | ✓ |
| 6th Class Pupil Questionnaire | ✓ | ✓ | ✓ |
| 6th Class Teacher Questionnaire | ✓ | ✓ | ✓ |
| 6th Class Parent/Guardian Questionnaire | ✓ | ✓ | - |
| School Principal Questionnaire | ✓ | ✓ | ✓ |

During the summer and autumn of 2020, all questionnaires were revised to acknowledge the extraordinary circumstances caused by the pandemic. Paper versions of the questionnaires were sent to all schools and school staff were informed that digital (online) versions of the teacher and school questionnaires were also available on request.

In previous cycles, Inspectors from the Department of Education oversaw the administration of testing in schools, delivered assessment materials to the schools on the test date and returned the materials to the ERC, to protect the security of the tests. However, in NAMER 2021, in the context of the pandemic, it was decided to change the administration procedure to minimise visits from external personnel to schools. Therefore, for this cycle, materials were sent directly to schools by post, and Inspectors visited a sample of schools (20%). Class teachers were assigned the role of test administrators and school coordinators acted as quality/security monitors. Detailed manuals were provided in which these roles were outlined and online training sessions were offered to class teachers, school coordinators and inspectors, during which opportunities were provided to seek clarification about their roles and the assessment procedures. The NAMER team members were also available throughout the testing period to provide support and clarification.

Overview of previous cycles of NAMER

The assessment frameworks developed and used for NAMER 2009 (Educational Research Centre, 2008, 2009) also underpinned the 2014 tests. In NAMER 2009 and 2014, pupils of selected Second and Sixth classes completed tests of English reading and Mathematics. There were four booklets at each grade level for English reading and four Second class mathematics booklets.

The administration of the Sixth class mathematics test in 2014 was taken as an opportunity to assess the feasibility of reducing the length of the mathematics test, following the expression of concern by class teachers that the 2009 test was very long (each of the booklets administered in 2009 contained three blocks of 25 test items). In 2014, five experimental, shortened versions of the mathematics test booklets were administered to pupils in a randomly selected 20% of participating schools that had pupils in Sixth class. These shortened booklets contained two blocks of 25 items, with a common block in the second position and one of each of the remaining blocks in the first position.

In 2014, significant improvements were observed in Second class pupils' performance on all reading content areas and process skills, compared with 2009¹. Targets that had been set in the National Strategy for Literacy and Numeracy (Department of Education & Skills, 2011) for achievement by 2020 were surpassed in 2014. Specifically, these targets were to increase the percentage of pupils achieving at the highest proficiency levels by at least 5 percentage points, and to reduce the percentage of pupils performing at the lowest proficiency levels by at least 5 percentage points. The results from NAMER 2014 showed a decrease of 13 percentage points in the percentage of pupils performing at Level 1 or below, and an 11-point increase in the percentage of pupils performing at Level 3 or above (Shiel et al., 2014).

Similarly, on the Sixth class mathematics assessment in 2014, pupils performed significantly higher than their counterparts in 2009 on all content areas and process skills. Pupils' overall mean mathematics score was 12 score points higher than in 2009. In terms of proficiency levels, the percentage of pupils performing at or below Level 1 decreased by 9 percentage points between 2009 and 2014, while the percentage of pupils performing at or above Level 3 increased by 7 percentage points. These results surpassed the targets set in the National Strategy for Literacy and Numeracy in 2011 (Shiel et al., 2014).

¹ Prior to NAMER 2014, with the exception of reading between 1972 and 1980, no national assessments had shown an increase in average performance. For example, there was no change observed in reading among Fifth class pupils between 1998 and 2004 (Eivers et al., 2005), and Mathematics in Fourth class between 1999 and 2004 remained stable (Shiel et al., 2006), despite the introduction of new curricula in 1999.

International assessments

Ireland's performance on international assessments since the last cycle of NAMER is outlined in this section. International assessments allow for comparison of achievement levels and trends across countries, but the links between test content and national curricula and/or policies may not be as strong as in a national assessment.

Since NAMER 2014, the following international assessments have been conducted: The Progress in International Reading Literacy Study (PIRLS) 2016, PIRLS 2021, the Trends in International Mathematics and Science Study (TIMSS) 2015, TIMSS 2019, the Programme for International Student Assessment (PISA) 2015 and PISA 2018. Although PISA is conducted at post-primary level, it is considered in this section alongside large-scale assessments conducted at primary level because patterns of performance on PISA may point towards issues of curriculum, teaching or learning at either level of the education system, or indeed both.

PIRLS 2016 and PIRLS 2021

PIRLS is a project of the International Association for the Evaluation of Educational Achievement (IEA) that assesses reading achievement among Fourth class pupils (Mullis & Martin, 2015). Since 2001 it has been conducted every five years, and Ireland participated for the first time in 2011. In PIRLS 2016, Ireland's mean score was 567, which was significantly above the study's centrepiece (500), and higher than all but two other participating countries (the Russian Federation and Singapore) (Eivers et al., 2017). Ireland's mean score in 2016 was 15 score points higher than its mean score in PIRLS 2011, and this was one of the largest increases amongst its comparison countries. Girls achieved a mean score of 572 in PIRLS 2016, which was statistically significantly higher than boys' mean score of 561. This gap of 12 score points was smaller than the international average gender difference of 19 points (which also favours girls). Twenty-one percent of pupils in Ireland performed at the advanced PIRLS benchmark (which indicates pupils' abilities to use more advanced reading skills). This compares favourably with the international median of 10% and it also represents a statistically significant increase on the 16% of pupils performing at the advanced benchmark in 2011.

In 2016, Ireland was one of fourteen countries to administer ePIRLS, which measures digital literacy, alongside the main PIRLS assessment. Ireland's mean score on ePIRLS was 567, which, like the mean score on the paper-based test, was significantly higher than the international average. Singapore was the only country with a significantly higher mean score on ePIRLS than Ireland (Eivers et al., 2017).

In PIRLS 2021, due to the disruption caused by the COVID-19 pandemic, some countries tested pupils at the end of Grade 4 (Fourth class in Ireland), while others, including Ireland, tested pupils at the start of Grade 5 (Fifth class). Another subset of countries tested pupils at the end of Grade 4, one year later than planned.

The mean age of pupils who completed the assessment in autumn at the start of Grade 5 (Fifth class) was, on average, 6 months older than the typical PIRLS Grade 4 (Fourth class) cohort, both in previous cycles in Ireland and internationally among countries who tested at the end of Grade 4 in PIRLS 2021. Pupils in Fifth class in Ireland were tested at a different point in the school year, soon after the summer break, compared with pupils who were tested at the end of Fourth class in previous cycles.

Pupils in Ireland achieved a mean reading score of 577 in PIRLS 2021, which is 11 points higher than Ireland's mean score in 2016. Due to certain caveats around the pandemic disruption, it is not possible to conclude that this represents an improvement in Irish pupils' mean reading performance since 2016. Delaney et al. (2023) propose that mean reading achievement in Ireland has remained relatively stable between 2016 and 2021. In terms of gender differences, girls in Ireland had a mean score that was 11 points higher than boys in PIRLS 2021; the magnitude of this gender gap is similar to that observed in 2016, and is comparatively small in relation to the international average.

TIMSS 2015 and TIMSS 2019

TIMSS is another assessment of the IEA that is conducted every four years and monitors trends in mathematics and science achievement (Mullis & Martin, 2017). At primary level, pupils in Fourth class are assessed. TIMSS was first carried out in 1995 and Ireland has participated on four occasions at primary level – in 1995, 2011, 2015 and 2019.

In the mathematics assessment of TIMSS 2015, pupils in Fourth class in Ireland achieved a mean score of 547, which was significantly above the international centrepiece. Pupils in seven countries (Singapore, Hong Kong, the Republic of Korea, Chinese Taipei, Japan, Northern Ireland and the Russian Federation) achieved significantly higher mathematics scores than pupils in Ireland, while pupils in Ireland achieved significantly higher scores than pupils in 37 other countries. Ireland's mathematics performance was not significantly different from Norway, England, Belgium (Flemish) and Kazakhstan. Boys' mean mathematics score was 4 points higher than girls' but the difference was not statistically significant. Fourth class pupils' mean mathematics achievement in 2015 was significantly higher than in 2011 (527) and 1995 (523) (Clerkin et al., 2016).

Performance in Mathematics remained stable in TIMSS 2019 - pupils in Fourth class in Ireland achieved a mean score of 548. This was statistically significantly higher than the score in 46 other countries and similar to the mean score in four countries (England, Norway, Latvia and Lithuania). Only seven countries had a significantly higher maths score than Ireland – these were Singapore, Chinese Taipei, Japan, Korea, the Russian Federation and Northern Ireland. There were no significant differences between boys and girls in Ireland in TIMSS 2019. More than half of Fourth class pupils (52 %) were classified as having reached the high mathematical benchmark, and 15 % reached the advanced benchmark, compared to an international median of 7 % (Perkins & Clerkin, 2020).

PISA 2015 and PISA 2018

PISA is an assessment of the Organisation for Economic Co-operation and Development (OECD) which measures 15-year olds' knowledge and skills in Reading literacy, Mathematics and Science (OECD, 2019a). In each cycle of PISA, one subject area is the major domain and the other two are minor domains. In more recent cycles of PISA, additional domains have been developed and made available for optional administration to students – for example, creative problem solving or global competence. Since 2015, most participating countries, including Ireland, have administered PISA by computer. At the time of writing, PISA 2022 is underway, and Mathematics is the major domain. In PISA 2015, Science was the major domain and in PISA 2018, Reading literacy was the major domain.

PISA – Reading literacy

In PISA 2015, Ireland's mean score on the reading scale was 520.8, which was significantly higher than the OECD average of 492.5. Ireland ranked 5th out of 70 participating countries on reading literacy and 3rd out of 35 OECD countries. Higher performing students in Ireland (i.e., those performing at the 90th percentile) had a mean score of 628.6, which was significantly higher than their peers in Northern Ireland (604.8) and significantly higher than students on average across OECD countries (612.7). However, higher performing students in Ireland performed significantly less well than students in Singapore (657.3), Canada (641.9), Finland (639.9), New Zealand (642.6), and France (636.9) (Shiel et al., 2016).

The PISA reading literacy scale has been divided by the OECD into seven proficiency levels, each of which represent different clusters of skills and the types of tasks that students at each level would be capable of completing successfully. Three processes underlie the PISA reading test items: 'access and retrieve', 'integrate and interpret', and 'reflect and evaluate'. In 2015, on average across OECD countries, about one-fifth of pupils performed at the lowest level of proficiency. In Ireland, 10.2 % of students performed at this lowest proficiency level, which is a similar percentage to students in Estonia, Canada, Finland and

Singapore (Shiel et al., 2016). According to the OECD (2016), students performing at this lowest level do not have sufficient reading skills to deal with future needs in further learning or real life. The proportion of students performing at the highest proficiency levels was 10.7 %, which is similar to Estonia, Germany, Sweden and the United States, but lower than Singapore, Canada, New Zealand or Finland. Ireland had a greater proportion of students performing at the highest proficiency level than the OECD average (8.3 %). Female students performed significantly higher than male students on the reading literacy scale in PISA 2015, by a difference of 12 score points (Shiel et al., 2016).

In PISA 2018, Ireland's mean score on the reading scale was 518.1, which was significantly higher than the OECD average of 487.1 (OECD, 2019b). Ireland was ranked 4th out of 36 OECD countries and 8th out of 77 participating countries/economies for which data were available. The only countries/economies that performed significantly higher than Ireland in reading literacy were Beijing, Shanghai, Jiangsu and Zhejiang (China), Singapore and Macao (China). In PISA 2018, the reading scale was divided into eight proficiency levels (additional lower levels of proficiency were included to allow for better description of the proficiency of students at the lower end of the proficiency distribution). In Ireland, 11.8 % of students performed at the lowest proficiency levels in reading, which was a statistically significantly lower proportion than the OECD average (McKeown et al., 2019). The proportion of students in Ireland performing at the highest proficiency levels was 12.1 %, which is higher than in Northern Ireland (9.5 %), not significantly different from the United States, Poland, New Zealand, Korea, and Sweden, but below the proportions in Singapore and Canada. The proportion of higher-achieving students in Ireland was statistically significantly higher than the OECD average (8.7 %). Female students in Ireland in PISA 2018 significantly outperformed male students by 23.2 score points, which was among the smallest gaps across comparison countries. On average across OECD countries, the gender difference in favour of females was 29.7 score points.

PISA – Mathematics

In PISA 2015, Ireland's mean score on the overall mathematics scale (503.7) was significantly higher than the OECD average (490.2). Ireland ranked 18th out of all 70 participating countries and 13th out of the 35 OECD countries. In terms of percentiles, students in Ireland had higher scores than the corresponding OECD averages at the 5th, 10th and 25th percentiles, but the scores of students in Ireland at the 75th, 90th and 95th percentiles were not significantly different.

The PISA mathematics scale is divided into six proficiency levels. In Ireland in 2015, 15 % of students performed below Level 2 (indicating weak performance), compared with the OECD average of 23.4 %. Only three other comparison countries (Japan, Estonia and Singapore) had fewer students than Ireland performing at this low level of proficiency. At the higher proficiency levels, 9.8 % of students in Ireland performed at Levels 5-6 (indicating strong performance), compared with the OECD average of 10.7 %. Male students in Ireland (511.6) outperformed female students (495.4) on PISA 2015 Mathematics by 16.1 points, which is a statistically significant difference. The gap between male and female students in Ireland is higher than the gap on average across OECD countries, where male students' mean score is 7.9 points higher than female students' score (Shiel et al., 2016).

In PISA 2018, Ireland's mean overall mathematics performance (499.6) was significantly above the OECD average of 489.3 and the EU average of 488.6. Ireland was ranked 16th out of 37 OECD countries and 21st out of the 78 participating countries/economies for which valid data were available. The proportion of pupils performing at the lowest proficiency levels (below Level 2) in Ireland was 15.7 %, compared with 24 % on average across OECD countries. The proportion of students in Ireland performing at the highest proficiency levels was 8.2 %, which is significantly lower than the average across OECD countries (10.9 %). Male students in Ireland achieved a mean score of 502.6 on PISA 2018 Mathematics, which was not statistically significantly different from female students' mean score of 496.7. The gender difference in Ireland (5.9 score points) is only slightly larger than the OECD average difference (5.2) in favour of males, although the latter is statistically significant (McKeown et al., 2019).

Interim Review of the National Literacy and Numeracy Strategy

The National Strategy for Literacy and Numeracy was launched in 2011 with an implementation period from 2011 to 2020 (Department of Education and Skills, 2011). Specific performance targets for the National Assessments were set in the Strategy, meaning that the National Assessments would play an important role in evaluating the impact of the implementation of the Strategy.

Specifically, the Strategy aimed to increase the proportion of pupils performing at the highest proficiency levels by at least 5 percentage points in both English reading and Mathematics at both Second and Sixth class, and to decrease the proportion of pupils performing at the lowest proficiency levels by at least 5 percentage points, in both subjects and at both grade levels (Department of Education and Skills, 2011). These targets were surpassed in NAMER 2014, and the following year, a review of the Strategy commenced, in order to observe and record successes and challenges approximately half-way through the implementation period. The review process involved a combination of questionnaires and follow-up meetings with various stakeholders (e.g., sections of the Department of Education, management bodies, teachers' unions, Education and Training Boards, the National Council for Curriculum and Assessment, and the National Parents' Council, among others). Focused meetings were also held with groups of principals, teachers and students (Department of Education and Skills, 2017b). New performance targets for the 2020 National Assessments were published in the Interim Review, with discrete performance targets for schools participating in the Delivering Equality of Opportunity in Schools (DEIS) programme, linked to the 2017 DEIS Plan (Department of Education and Skills, 2017a).

Table 1.2. Performance targets for English reading and Mathematics and results from NAMER 2009 and 2014

| | | NAMER 2009 results | Target for 2020 set in 2011 Strategy | NAMER 2014 results | New target for 2020 |
|---|-----------|-----------------------|---|-----------------------|------------------------|
| Reading: At or above Level 3 | 2nd Class | 35.0 % | 40.0 % | 45.5 % | 50.0 % |
| | 6th Class | 35.0 % | 40.0 % | 43.9 % | 50.0 % |
| Reading: At or below Level 1 | 2nd Class | 35.0 % | 30.0 % | 21.6 % | 20.0 % |
| | 6th Class | 35.0 % | 30.0 % | 24.8 % | 20.0 % |
| Mathematics: At or above Level 3 | 2nd Class | 35.0 % | 40.0 % | 47.2 % | 53.0 % |
| | 6th Class | 35.0 % | 40.0 % | 42.1 % | 50.0 % |
| Mathematics: At or below Level 1 | 2nd Class | 35.0 % | 30.0 % | 25.6 % | 20.0 % |
| | 6th Class | 35.0 % | 30.0 % | 26.5 % | 20.0 % |

CHAPTER 2

Assessment Frameworks

In this chapter, the assessment frameworks and the process of revising the assessment instruments for pupils in Second and Sixth classes in NAMER 2021 are outlined. The framework underpinning the English reading assessment is described first, followed by the mathematics assessment framework. The 2021 framework for English reading builds on the 2009 framework (also used in NAMER 2014 without any changes) and references the new Primary Language Curriculum/Curaclam Teanga na Bunscoile (PLC/CTB) (Department of Education and Skills, 2019). The mathematics assessment framework for NAMER 2021 is the same as the one used in 2009 and 2014.

Although these frameworks were written for the development of the assessments for Second and Sixth classes, due to the adaptations necessitated by the COVID-19 pandemic, only the Second class English reading assessment and the Sixth-class mathematics assessment were administered in 2021.

English reading assessment framework

Introduction

The English reading element of NAMER 2021 is the ninth in a series of national assessments dating back to 1972. These assessments measure pupils' achievement in reading, based on the curriculum for primary schools (Department of Education and Science & National Council for Curriculum and Assessment, 1999a; Department of Education and Skills, 2019; National Council for Curriculum and Assessment, 2019), and gather data related to reading achievement.

The English reading assessment has focused on a range of grade levels since 1972.³ Since 2009 pupils have been assessed at the end of Second and Sixth classes, to coincide with the end of the Junior and Senior cycles at primary level. The most recent National Assessments involving English reading and Mathematics took place in 2014.

The 2021 NAMER assessment framework for English reading builds upon the framework that was used in 2009 and 2014, while repositioning some elements in light of the revised PLC for junior classes (National Council for Curriculum and Assessment, 2019). The assessment takes place in the broader context of Ireland's continued involvement in PIRLS at Fourth class, for which there is a separate framework (Mullis & Martin, 2015).

Defining reading literacy/comprehension

In recent years, there has been a tendency to define literacy more broadly than as reading or reading literacy. For example, literacy is defined by the National Literacy and Numeracy Strategy as "the capacity to read, understand and critically appreciate various forms of communication including spoken language, printed text, broadcast media, and digital media" (Department of Education and Skills, 2011, p. 8).

The PLC does not explicitly define reading comprehension (Department of Education and Skills, 2019). However, reading – encompassing decoding, fluency, word meanings, comprehension, engagement and motivation – is one of the three integrated strands of language outlined in the curriculum document. The other two strands are oral language and writing. The aims of the PLC include supporting teachers to "encourage children to engage personally with and think critically about a broad range of spoken, gesticulated, written and multimodal texts" and to "promote a positive disposition towards communication and language by fostering within children a lifelong interest in and love of language learning for personal

³ From 1972 to 2004, pupils participating in the NAER were assessed at Fourth and Fifth class levels. This changed in 2004 when the NAER moved to an assessment of achievement at First and Fifth class levels. In 2009, a decision was made by the Department of Education and Skills to assess pupils at the end of Second and Sixth classes in order to coincide with the end of cycles in the primary school curriculum.

enjoyment and enrichment” (National Council for Curriculum and Assessment, 2019, p. 27). The curriculum views skilled reading comprehension as being important for meaningful communication, and for children’s ability to participate in and learn from their environment.

Previous NAMER reading frameworks have drawn on the definitions used by international assessments such as PIRLS and PISA, which situate reading comprehension within reading literacy. The 2004 and 2009 NAMER frameworks defined reading literacy (incorporating comprehension) as:

The process of constructing meaning through the dynamic interaction among the reader’s existing knowledge, the information suggested by the written language, and the context of the reading situation. Young readers read to learn, to participate in communities of readers, and for enjoyment (Eivers et al., 2005, p. 15).

The following definition is used for NAMER 2021:

Reading literacy enables the reader to engage with, comprehend, use, and respond to written language in varied forms. A reader can construct and extend meaning through the dynamic interactions between their existing knowledge, the information suggested by the written language, and the context of the reading situation.

NAMER aims to assess reading literacy through the assessment of reading comprehension and vocabulary knowledge. While retaining the essential components of the definition used in previous National Assessments, the current definition emphasises engagement with, use of and response to written language in varied forms. The definition is sufficiently flexible to incorporate comprehension of digital texts, if it is decided that these will feature in future NAMER cycles.

Incorporating revisions to the syllabus into the framework

The English reading framework for NAMER 2021 draws on the learning outcomes, progression milestones and progress steps in the PLC.

Among the key learning outcomes for reading in the PLC that are relevant for NAMER 2021 (Second class reading) are:

- Use conventions of print and sentence structure to help understand text.
- Draw on a developing knowledge of word structure and word parts, including some prefixes and suffixes, to clarify and discuss the meaning of words.
- Use phonic knowledge and a range of word identification strategies with flexibility and confidence when reading instructional and independent-level texts.
- Acquire appropriate vocabulary to support the independent comprehension of text.
- Experience and respond to the aesthetic, creative and imaginative aspects of texts and a wider range of genres justifying preferences and opinions.
- Draw on a repertoire of comprehension strategies and background knowledge to comprehend text independently or collaboratively.
- Recall, discuss and sequence significant details and identify key points of information in text.
- Read instructional and independent-level texts in a range of genres with reasonable pace and understanding, and self-correct independently.

Vocabulary and comprehension strategies referred to in the PLC include the following Progression Milestones (National Council for Curriculum and Assessment, 2019, p. 33), though it is recognised that these strategies are on a continuum and will not have been achieved by all children by the end of Second class:

- The child reads a range of high-frequency and CVC (consonant, vowel, consonant) words and uses some contextual as well as pictorial cues to aid comprehension. (Progression milestone E)
- He/she modifies predictions based on new information and shows understanding through discussion of texts read. (Progression milestone E)
- The child enjoys listening to, reading and exploring text showing knowledge of the structure of a range of narrative and procedural text genres. (Progression milestone F)
- He/she understands messages conveyed in illustrations and uses sight words, rhyming words, syllables, common letter patterns and text cues to read. (Progression milestone F)
- The child reads familiar texts independently using full stops to punctuate. (Progression milestone F)
- He/she begins to monitor comprehension, asks questions about texts and offers some reasons for predictions. (Progression milestone F)
- The child enjoys choosing and exploring texts for specific purposes and generates and responds to questions on a range of text genres providing justifications for responses. (Progression milestone G)
- He/she reads a range of words with digraphs, letter patterns and syllables and uses prior knowledge, contextual cues, some punctuation and adds some intonation while reading aloud. (Progression milestone G)
- The child locates key information using text features such as table of contents and sub-headings and begins to use dictionaries for word meanings. (Progression milestone G)
- He/she discusses miscomprehension in text (Progression milestone G)
- The child enjoys a range of genres and reads for specific purposes, justifying preferences and opinions. (Progression milestone H)
- He/she uses a range of comprehension strategies and text features such as indexes to help understand texts, locate key information and disregard unnecessary detail, and cites specific evidence in text to support a viewpoint. (Progression milestone H)
- The child uses a range of word identification strategies and knowledge of synonyms, homonyms, antonyms, affixes and root words to understand unfamiliar words (Progression milestone H)

Reading vocabulary and comprehension in the framework

In this framework, vocabulary is identified as a foundational skill, and is assessed due to its close bidirectional relationship with reading comprehension. Other skills identified as foundational to reading comprehension include aspects of decoding such as phonemic awareness, orthographic knowledge, word knowledge and oral reading fluency. These skills are not assessed in NACER 2021 as most children are expected to have mastered these skills by the end of Second class. Rather, this assessment focuses on measuring comprehension.

The role of vocabulary in reading comprehension is complex. A large body of research suggests that vocabulary knowledge and skilled reading comprehension are closely and positively linked and share the same underlying processes (Biemiller, 2003; Cain et al., 2004; Currie & Cain, 2015; Hart & Risley, 1995; Language and Reading Research Consortium, 2015; Oakhill & Cain, 2012). For this reason, a vocabulary assessment forms part of the NAMER English reading test along with the passages and questions which assess reading comprehension. As in 2009 and 2014, the same sets of vocabulary items are used across multiple forms or booklets in NAMER 2021 to facilitate the placement of several test forms onto a single underlying scale.

The relationship between reading comprehension and vocabulary knowledge is well known (Shany & Biemiller, 2010; Verhoeven & van Leeuwe, 2008), but is not necessarily well understood. Studies suggest a bidirectional relationship (M. G. McKeown et al., 1983), which becomes more important as a reader becomes more skilled (Kame'enui & Baumann, 2012). The bidirectional nature of the relationship suggests that a wide vocabulary reflects knowledge of individual words but also supports the processing of unfamiliar words through strategic inferencing, for example analogy (Ehri, 2014). The value of continuing to include vocabulary as an element of NAMER reading assessment is reinforced by the designation of vocabulary as a learning outcome under 'Understanding the Content and Structure of Language' in the PLC.

Specific elements of the NAMER 2021 framework for English reading

Reading Purposes

In keeping with the framework used for the 2009 and 2014 assessments, two reading purposes are used for classifying texts in the 2021 assessment. The reading purposes are described as (i) reading for literary experience and (ii) reading to acquire and use information.

Reading for literary experience. According to Mullis et al. (2015, p. 16), “readers engage with the text to become involved in events, settings, actions, consequences, characters, atmosphere, feelings, and ideas, and to enjoy language itself. In order to understand and appreciate literature, each reader must bring to the text his or her own experiences, feelings, appreciation of language, and knowledge of literary forms”.

Reading to acquire and use information. Mullis et al. (2006, p. 19) describe this as: “the reader engages ... with aspects of the real universe. Through informational texts, one can understand how the world is and has been, and why things work as they do. Readers can go beyond the acquisition of information and use it in learning and in action”. Informational texts may be ordered (a) chronologically, (b) logically, (c) in expository form, or (d) as non-continuous text or in some combination of these. These text types are further elaborated on as:

- a) Chronological presentations generally involve recounting of events (e.g., historical facts, biographies) or presentation of procedures (e.g., recipes).
- b) Logical presentations involve presentation of ideas in an orderly sequence (e.g., text comparing and contrasting weather conditions, persuasive arguments).
- c) Expository texts present explanations or describe people, events, or things.
- d) Non-continuous texts may be presented in the form of lists, diagrams, charts, graphs, timetables, advertisements (Mullis et al., 2006).

These elements of the NAMER 2021 (and earlier) frameworks can be aligned with the PLC (Junior Infants to Second class) (Table 2.1).

Table 2.1. Alignment between text types (Reading purposes) in the NAMER 2021 framework and current curricula at primary level

| Text Type | Example | NAMER 2021 | PLC – Second |
|---|-----------------------------------|------------|--------------|
| Reading for Literary experience | | | |
| Narrative | Short story | ✓ | ✓ |
| Recount | Retelling of story | ✓ | ✓ |
| Drama | Excerpt from a play | ✓ | ✓* |
| Poetry | Short poems | ✓ | ✓ |
| Reading to acquire and use information | | | |
| Informational /Expository | | | |
| Expository | Newspaper or magazine article | ✓ | ✓* |
| Procedure | Recipes, Make and Do instructions | ✓ | ✓ |
| Chronological Account | Biography | ✓ | ✓ |

*Text type is implied (e.g., drama in oral language)

Table 2.1 suggests a good level of alignment between the NAMER 2021 framework and current curriculum documents, including the PLC, which refers to pupil engagement with “a range of narrative and procedural text genres” (National Council for Curriculum and Assessment, 2019, p. 67). Further refinement of the purposes for reading in the NAMER 2021 framework may be required in the future if digital texts are assessed in addition to paper-based ones.

In order to better reflect the reading experiences of children under the new PLC, such texts as adventure texts, historical fiction, drama, fantasy, science fiction and biographical accounts were considered for inclusion among the new texts to be developed.

Reading Processes

The reading processes to be assessed in NAMER 2021 reflect those assessed in PIRLS and the 2009 and 2014 National Assessments.

Four processes are specified: Retrieval, Straightforward Inferential Processing; Interpretation and Integration; and Examination and Evaluation.

Retrieval requires the reader to decode a text accurately, to understand what is stated in the text, and how it relates to the information that is sought; e.g., looking for specific information, events, ideas, definitions or phrases; identifying the setting of a story (e.g., time, place); finding the main theme of a text when explicitly stated.

Straightforward inferential processing requires direct inference from a text; e.g., deducing or inferring that one event caused another; determining the main point of a series of arguments; identifying generalisations in a text; describing the relationships between two characters.

Interpretation and integration requires some integration of personal knowledge or experience with meaning that resides in a text to construct a more complete understanding of the text; e.g., discerning the overall message or theme of a text; considering an alternative to actions of characters; comparing and contrasting text information; inferring the mood or tone of a story; or applying text information to a real-world situation.

Examination and evaluation involves evaluation of a text, either from a personal perspective or a more critical and objective viewpoint; e.g., evaluating the plausibility of events described in a text; identifying and commenting on the structure and organisation of texts, including grammatical and presentational

features; judging the completeness or clarity of information in a text; determining an author’s perspective on the central topic in a text; identifying and commenting on the writer’s purposes and viewpoints, and the overall effect of the text on the reader; and relating texts to their social, cultural, and historical contexts and literary traditions.

The Reading processes assessed in NAMER are broadly aligned with those in the PLC (Second class) (Table 2.2).

Table 2.2. Alignment between Reading Processes in the NAMER 2021 Framework and Current Curriculum at Primary Level

| Reading Process | Example | NAMER 2021 | PLC – Second |
|----------------------------------|---|------------|--------------|
| Retrieve | Recall information explicitly stated in the text | ✓ | ✓ |
| Infer | Make a straightforward inference in the text | ✓ | ✓ |
| Interpret & Integrate | Integrate information in the text with personal knowledge to create a more complete understanding | ✓ | ✓ |
| Examine and evaluate | Identify and comment on the structure and organisation of texts | ✓ | ✓ |

Correlations between reading subscales

Correlation coefficients can indicate how strongly the scores on different aspects of the assessment are related to each other. The correlation between the Vocabulary and the Comprehension subscales is equal to .77. Table 2.3 presents the correlations among the three reading process scales. The strongest correlation is between the Retrieve and Infer subscales ($r = .77$). The correlations between the Retrieve and Interpret & Integrate subscales ($r = .70$) and the correlation between the Infer and Interpret & Integrate subscales ($r = .71$) are of similar strength. All three correlations among the three subscales are considered strong.

Table 2.3. Correlations between reading process subscale scores

| | Retrieve | Infer | Interpret & Integrate |
|----------------------------------|------------|------------|-----------------------|
| Retrieve | - | - | - |
| Infer | .77 | - | - |
| Interpret & Integrate | .70 | .71 | - |

Note. Correlation coefficients in bold indicate statistically significant correlations. EAP rather than PV scores were used to estimate the correlation coefficients among the subscales.

Structure of the Tests

In designing the tests for 2021, consideration was given to previous iterations of the National Assessments (especially 2009 and 2014), to facilitate reporting on trends in performance.

At Second class, only multiple-choice items were presented, as many pupils at this class level might not have developed sufficient writing/spelling skills to demonstrate their knowledge on constructed response items. In a departure from previous National Assessments, NAMER 2021 included a small number of items assessing *Examine and evaluate*, using the multiple-choice format. As noted above, these processes are included in the new PLC (Junior Infants to Second class), and it was considered worthwhile to capture some information about pupils’ abilities in this sphere.

Four separate forms were developed for the Second class reading assessment. Each form consists of 20 vocabulary items, followed by five reading comprehension texts of varying length and associated multiple-choice items. Total administration time for this level was approximately 90 minutes (total testing time, excluding set up and breaks, is approximately 1 hour and 10 minutes). Table 2.4 below provides a broad outline of the structure of the tests.

Table 2.4. Table of Specifications – Numbers of Items by Reading Purpose and Process – Second Class

| Processes | Purposes | | | |
|---|---------------------------------|------------|--|------------|
| | Reading for literary experience | | Reading to acquire and use information | |
| | N of items | % of items | N of items | % of items |
| Focus on and retrieve explicitly stated information | 37 | 58 % | 20 | 29 % |
| Make straightforward inferences | 16 | 25 % | 26 | 38 % |
| Interpret and integrate ideas and information | 11 | 17 % | 21 | 30 % |
| Examine and evaluate content | 0 | 0 % | 2 | 3 % |
| Core Test Subtotal | 64 | | 69 | |
| Foundation Component (Vocabulary) | 20 | | | |

Assessment development

Passage selection

A key component of national assessments is to monitor progress over time. In order to do this, there must be strong continuity from cycle to cycle in terms of the test framework and the passages and items used. However, it is still possible, and desirable, to change some test items from one cycle to the next. Changes may be needed if test items no longer operate as expected (for example, if they no longer discriminate among pupils) or simply become dated.

A large number of reading comprehension passages and vocabulary items were retained from previous National Assessments, allowing for continuity of assessment and for trend analysis. However, a selection of passages was retired after the 2014 assessment, and these were replaced in the development phase of NAMER 2021.

A team of experienced teachers, literacy experts, and researchers were employed as item writers in preparation for the Pilot Study in 2019. A wide variety of suitable, engaging texts for Second and Sixth classes were selected, and items appropriate to the texts selected were developed.

Texts were chosen based on the passage selection criteria utilised in the development of the new Drumcondra Reading Tests (Educational Research Centre, 2019). These included issues around the length and difficulty of the passage, the level of likely engagement, the verisimilitude of the text, as well as considerations of diversity and non-bias within the text.

In choosing texts, NAMER 2021 guidelines say that texts should:

- be engaging to a wide variety of pupils from a range of backgrounds
- contain themes and situations represented in the PLC and the Primary Curriculum – English, and encountered in real life
- be well-written, and coherent within the confines of the extract
- represent a variety of genres and styles
- be of an appropriate complexity and difficulty

Length and difficulty

The National Assessments contain texts of varying length at each grade level. In NAMER 2021, Second Class texts were between 82 words and 356 words long.

Texts are pitched at average-to-high difficulty for the target grade level. Points of reference for determining appropriate difficulty include the vocabulary and syntax used in extracts in popular textbooks, lists of most popular authors and books for various age and class groups. Test developers considered a number of factors when evaluating the difficulty of a text: its coherence at macro- and micro-levels; the extent to which bridging inferences are required; sentence length and syntactic complexity; and the accessibility of subject matter and form.

Non-bias and diversity

Text context and content should not favour any subgroup of pupils over another in terms of prior knowledge or probable engagement. For instance, girls and boys are likely to respond with different levels of knowledge and engagement to a review of a popular children’s film which has been marketed almost exclusively to girls. A text which assumes knowledge of the rituals of one religion (e.g., First Communion within a Catholic tradition) is likely to work differently for children who have grown up within that religion, and those who have not.

Texts were selected for the reading test with due consideration to non-bias and diversity. An effort has been made to reference a wide variety of diverse cultures and diverse family compositions, in recognition of the increased diversity within Irish schools and of the curricular aim that language-learning should help foster an appreciation of diversity (Department of Education and Science & National Council for Curriculum and Assessment, 1999a; National Council for Curriculum and Assessment, 2019). Moreover, certain groups of children are chronically under-represented in mainstream discourse, including in children’s books. These include Traveller children, children from migrant families, and children with disabilities.

The passages within each form were selected to provide internal contrast in terms of subject matter, style, and genre, to promote engagement for as many pupils as possible. Thus, a child who disliked the subject and style of Passage A may have had a “second chance” to engage with the test when they reached the contrasting Passage B.

Further considerations

Passages selected for inclusion in the test should be unfamiliar to most pupils. Therefore, extracts from very popular children’s books – or books likely to become extremely popular within the lifetime of the test – were considered unsuitable.

Passages were also screened for their potential to date quickly. For instance, a passage about a messaging app that is currently widely used may become obsolete in a few years as technology moves on.

In 2021, at Second class, 15 new vocabulary items and 15 reading passages comprised of 145 items (distributed over five test booklets) were piloted. A number of these new items replaced retired 2014 items in the final NAMER 2021 assessment (Table 2.5).

Table 2.5. Number of trend and new items in NAMER 2021 by content area, Second Class English Reading

| | N of trend items | N of new items | Total |
|---------------|------------------|----------------|------------|
| Vocabulary | 15 | 5 | 20 |
| Comprehension | 95 | 38 | 133 |
| Total | 110 | 43 | 153 |

The use of Item Response Theory (IRT) scaling facilitates the process of dropping and adding items across cycles while maintaining trends. In Second class English reading, three trend items appeared to perform differently in 2021 compared to 2014; these were treated as new items in 2021, while one more item was excluded from the 2021 data as it did not function correctly.

Summary

The framework underpinning the NAMER English reading assessment underwent significant development in 2009 in line with the decision to focus the assessment on Sixth and Second class. No changes were made to the framework in 2014, though a small number of passages and associated items were dropped and replaced.

Some minor alterations to the assessment framework have been implemented for 2021 to reflect changes in the PLC. Like the PLC, the 2021 framework also seeks to take into account changing demographics in Ireland, including increased numbers of pupils who speak languages other than English or Gaeilge at home.

Scores on all of the reading subscales are strongly correlated.

Mathematics assessment framework

The NAMER 2021 assessment framework for Mathematics is the same as that used in the previous cycles of the National Assessments. While a brief overview is provided here, a more detailed de-scription of the framework and the test development process can be found in Educational Research Centre (2009) and in Eivers et al. (2010).

The mathematics framework drew directly on the definition of Mathematics in the Primary School Mathematics Curriculum (PSMC) which sees Mathematics as:

...the science of magnitude, number, shape, space, and their relationships and also as a universal language based on symbols and diagrams. It involves the handling (arrangement, analysis, manipulation and communication) of information, the making of predictions and the solving of problems through the use of language that is both concise and accurate. (Department of Education and Science & National Council for Curriculum and Assessment, 1999b, p. 2).

The PSMC contains instructional objectives for each class level which are formed from a combination of two main dimensions: mathematical content strands and cognitive process skills. The content strands of

⁴ Number and Algebra are combined in the NAMER assessment framework as there was an insufficient number of Algebra items for them to be scaled separately from Number.

the PSMC are Number, Algebra⁴, Shape & Space, Measures, and Data. The cognitive process skills are Understanding & Recalling, Implementing, Reasoning, Connecting & Integrating, Applying & Problem Solving, and Communicating & Expressing. The Professional Development Service for Teachers (PDST) provides documentation that summarises the key elements of each process skill across primary grade levels.⁵ Each test item is classified by content strand and process skill. Items assessing the Communicating & Expressing skill could not be included in the National Assessments, given the pencil-and-paper format of the tests.

The NAMER mathematics tests are designed so that the distributions of items assessing the content strands and cognitive processes closely approximate the distribution of these elements in the PSMC objectives (see Tables 2.6 and 2.7 for the distributions at Sixth class in NAMER 2009, 2014 and 2021).⁶ As the item pool for Data is small, particular care should be exercised in drawing inferences about performance on this content strand.

Table 2.6. Classification of final mathematics items by content strand, Sixth class

| | NAMER 2009 | | NAMER 2014 | | NAMER 2021 | | % PSMC Objectives |
|------------------|------------|--------------|------------|--------------|------------|--------------|-------------------|
| | N of items | % of items | N of items | % of items | N of items | % of items | |
| Number & Algebra | 69 | 46.0 | 69 | 46.0 | 69 | 46.0 | 43.0 |
| Shape & Space | 32 | 21.3 | 32 | 21.3 | 32 | 21.3 | 21.0 |
| Measures | 31 | 20.7 | 31 | 20.7 | 31 | 20.7 | 24.0 |
| Data | 18 | 12.0 | 18 | 12.0 | 18 | 12.0 | 12.0 |
| Total | 150 | 100.0 | 150 | 100.0 | 150 | 100.0 | 100.0 |

Table 2.7. Classification of final mathematics items by process skill, Sixth class

| | NAMER 2009 | | NAMER 2014 | | NAMER 2021 | |
|-----------------------|------------|--------------|------------|--------------|------------|--------------|
| | N of items | % of items | N of items | % of items | N of items | % of items |
| Understand & Recall | 15 | 10.0 | 15 | 10.0 | 16 | 10.7 |
| Implement | 30 | 20.0 | 30 | 20.0 | 30 | 20.0 |
| Integrate & Connect | 8 | 5.3 | 8 | 5.3 | 8 | 5.3 |
| Reason | 47 | 31.3 | 47 | 31.3 | 47 | 31.3 |
| Apply & Problem-Solve | 50 | 33.3 | 50 | 33.3 | 49 | 32.7 |
| Total | 150 | 100.0 | 150 | 100.0 | 150 | 100.0 |

During the administration of NAMER 2009, some concern was expressed by several class teachers over the length of the Sixth class mathematics test (130 minutes in duration, compared to 90 minutes at Second class Mathematics, and 90 minutes for reading at both grade levels), and questions were raised about whether the test was unduly burdensome on pupils. In response to this feedback, the 2014 assessment was used as an opportunity to explore the feasibility of reducing the length of the Sixth class test booklets in future National Assessments.

⁵ https://pdst.ie/sites/default/files/Maths%20skill%20line%20of%20development_0.pdf

⁶ Due to an error that took place during the scaling of NAMER 2021 6th class mathematics data, two “Data” and one “Number/Algebra” items were misclassified to incorrect subscales. Analysis was done to confirm that this error did not have a significant impact on the subscale scores.

To this end, shortened versions of the mathematics test forms were administered to pupils in a random set of 20 % of participating English-medium schools that had Sixth class pupils. Five versions of these experimental, shortened booklets were administered. Each test booklet contained two sections, with a common block in the second position and one of each of the remaining blocks in the first position. Pupils who received the shortened test booklets had a mean scale score that was 4 points higher than pupils who took the regular-length test. For the purposes of the NAMER 2014 report, which looked at trends in achievement from 2009 to 2014, the scores for pupils taking the experimental booklet were excluded from the analysis; the sample of pupils and schools was sufficiently large to allow for the exclusion of these pupils without compromising estimates of statistics.

In NAMER 2021, a decision was made to administer the shortened version of the Sixth class mathematics test. Similarly to the NAMER 2014 shortened mathematics tests, each pupil answered two blocks of questions; 50 items in total. However, the layout of the 2021 booklets was different to the one used in 2014, so that all pupils answered a block without the use of calculators first, before moving to the second block where pupils are allowed to use a calculator. The linking of the results from 2021 and 2014 was achieved using all of the available long booklets from 2014, but only two of the six short booklets; four further short booklets consisting of two calculator blocks were excluded from scaling as the second block was easier when it followed another calculator block.

It should be noted that the transition to shorter mathematics booklets reduced the amount of data collected for each item. Therefore, mathematics subscale scores should be analysed and interpreted with caution, especially those that include small numbers of items, such as Connect, Understand/Recall and Data.

A key component of National Assessments is to monitor progress over time. In order to do this, there must be strong continuity from cycle to cycle in terms of the test framework and the items used. However, it is still possible, and desirable, to change some test items from one cycle to the next. Changes may be needed if test items no longer operate as expected (for example, if they no longer discriminate among pupils) or simply become dated. In 2021, a small number of new items replaced retired 2014 items in Sixth class (Table 2.8).

Table 2.8. Number of trend and new items in NAMER 2021 by content area, Sixth class Mathematics

| | N of trend items | N of new items | Total |
|-----------------------------|------------------|----------------|-------|
| Number & Algebra | 66 | 3 | 69 |
| Shape & Space | 29 | 3 | 32 |
| Measures | 30 | 1 | 31 |
| Data | 16 | 2 | 18 |
| Total | 141 | 9 | 150 |

The use of IRT scaling facilitates the process of dropping and adding items across cycles, while maintaining trends. In Sixth class Mathematics, one item was excluded from the analysis for the 2021 cycle because it performed significantly differently compared to the previous NAMER cycle.

Correlations between mathematics content subscales

Table 2.9 shows the correlations between the mathematics content areas in NAMER 2021. The strongest correlation is between the Number & Algebra and Measures subscales ($r = .76$). The correlation between Data and Number & Algebra is $.60$. The correlation between the Shape & Space and Number & Algebra subscales is $.58$, while the correlation between the Shape & Space and Measures subscales is $.56$. The correlation between Data and Measures is also $.56$. These five correlations are considered strong. A weaker correlation exists between Data and Shape & Space ($r = .46$).

Table 2.9. Correlations between mathematics content subscales

| | Number & Algebra | Shape & Space | Measures | Data |
|------------------|------------------|---------------|------------|------|
| Number & Algebra | -- | | | |
| Shape & Space | .58 | -- | | |
| Measures | .76 | .56 | -- | |
| Data | .60 | .46 | .56 | -- |

Note. Correlation coefficients in bold indicate statistically significant correlations. EAP rather than PV scores were used to estimate the correlation coefficients among the subscales.

Table 2.10 displays the correlations between the scores on the process subscales. The strongest correlation is between Reason and Apply & Problem Solve ($.77$), while the weakest correlation is between Understand & Recall and Integrate & Connect ($.49$). All of the correlations shown below are statistically significant.

Table 2.10 Correlations between mathematics process subscales

| | Understand & Recall | Implement | Reason | Integrate & Connect |
|-----------------------|---------------------|------------|------------|---------------------|
| Understand & Recall | - | - | - | - |
| Implement | .69 | - | - | - |
| Reason | .70 | .74 | - | - |
| Integrate & Connect | .49 | .49 | .54 | - |
| Apply & Problem Solve | .71 | .76 | .77 | .56 |

Note. Correlation coefficients in bold indicate statistically significant correlations. EAP rather than PV scores were used to estimate the correlation coefficients among the subscales.

CHAPTER 3

Methods, Procedures and Technical Notes

Pilot study

A pilot study for NAMER 2021 was conducted in April-June 2019. One of the main aims of the pilot was to try out and select new English reading and mathematics items to replace those which had been retired from older test booklets, as well as to trial new questionnaire items. For Second class reading, 15 new vocabulary items and 15 reading passages comprised of 145 items (distributed over five test booklets) were piloted. For Sixth class Mathematics, two booklets were piloted, each containing 26 items. As well as the test items, a large number of pupil, teacher and principal questionnaire items were administered and tested in the pilot study. Second class Mathematics, Sixth class English reading and parent/guardian questionnaire items at both class levels were also administered during the 2019 pilot study but these instruments were not administered in the main study due to the COVID-19 pandemic and the adaptations to NAMER 2021 described in Chapter 1.

The pilot test booklets and questionnaires were administered to a sample of Second class and Sixth class pupils in 40 vertical schools that had English as the primary language of instruction. To reduce the amount of work for younger pupils, Second class pupils took only one test (English reading or Mathematics) and they completed either the pupil questionnaire or their families were asked to complete the parent/guardian questionnaire. At Sixth class, pupils sat both assessments and both the pupil and the parent/guardian questionnaires were administered.

In NAMER 2021, five Vocabulary items and six reading passages with associated items were replaced at Second class with texts and items that had been trialled in the pilot study. At Sixth class, nine mathematics items were replaced. Replacement items at both grades were selected from the available pool of piloted items and aimed to maintain the difficulty of the test and ensure that new items examined similar content and processes to the retired items.

Main study sample design and administration

Sample design

The sample for NAMER 2021 was selected in two stages; first, schools were selected; then intact classes were selected within sampled schools. The target population consisted of all Second and Sixth class pupils in mainstream classes in primary schools in Ireland in the Department of Education 2018/19 list of schools; private schools and special schools were excluded. The sample was originally selected in 2019 for the planned administration of NAMER in spring 2020. As the study was postponed until spring 2021 due to the COVID-19 pandemic, the sample was carried over to 2021.

In NAMER 2021, DEIS Band 1 and Band 2 schools were oversampled in order to improve the accuracy of estimates for pupils attending such schools. In order to facilitate this, the sampling frame was split into several explicit strata based on DEIS category, size and whether the school was junior, senior or vertical. Within each explicit stratum probability proportional to size (PPS), systematic sampling was used with implicit stratification (sorting) by Gaeltacht type, school gender mix, and school size at grade. Two replacement schools were identified as part of this process with the 1st replacement being the school immediately after the selected school in the sorted list and the 2nd replacement being the school immediately prior to the selected school. In total, 195 schools were selected to participate in the study. Out of the 195 selected schools, seven schools did not participate in NAMER for various reasons; many related to the pandemic (e.g., high numbers of COVID-19 cases in the school). Table 3.1 presents the final school sample by school type and DEIS category.

Table 3.1. Number of participating schools by school type and DEIS category

| | Non-DEIS | DEIS Band 1 | DEIS Band 2 | DEIS Rural | Total |
|--------------|-----------|-------------|-------------|------------|------------|
| Junior | 8 | 9 | 4 | 1 | 22 |
| Senior | 7 | 9 | 4 | 1 | 21 |
| Vertical | 80 | 40 | 22 | 3 | 145 |
| Total | 95 | 58 | 30 | 5 | 188 |

The second stage of sampling was at the class level. The Department of Education supplied the ERC with details of the Second and Sixth classes in participating schools, excluding special classes. For each school, ERC staff randomly selected up to two intact classes at each grade level. In practice, this meant that in small- and medium-sized schools (i.e., those schools that had only one or two classes at each grade level), all pupils at the target grade level were selected. Pupils could be exempted at this stage if their teacher felt that it was appropriate to do so. The main reasons for exemption were limited proficiency in English, or certain learning and physical disabilities. However, it was emphasised to class teachers, school coordinators, principal teachers and inspectors that any pupil who could complete the assessment should be encouraged to do so, and that exemptions should be rare.

Response rates

Despite the challenges posed by the COVID-19 pandemic, within-school response rates in NAMER 2021 were high. Tables 3.2 and 3.3 present the response rates for Second and Sixth class in the last three cycles of the National Assessments. In Second class reading in NAMER 2021, around two percent of pupils were exempted, while this figure was lower in Sixth class. A further six to seven percent of Second and Sixth class pupils were absent on the testing day. Test data were collected for 91.7% of selected Second class pupils for English reading and for 91.4% of selected Sixth class pupils for Mathematics in NAMER 2021. The overall response rates for NAMER 2021 were slightly lower than 2014, but similar to 2009.⁷

Table 3.2. Response rates for the National Assessments, Second class reading

| | 2009 | | 2014 | | 2021 | |
|-----------------------|-------------|--------------|-------------|--------------|-------------|--------------|
| | <i>N</i> | % | <i>N</i> | % | <i>N</i> | % |
| Exempt | 53 | 1.3 | 48 | 1.1 | 104 | 1.8 |
| Absent | 307 | 7.3 | 223 | 5.1 | 365 | 6.4 |
| Tested | 3839 | 91.4 | 4099 | 93.8 | 5201 | 91.7 |
| Total enrolled | 4199 | 100.0 | 4370 | 100.0 | 5670 | 100.0 |

⁷ In Tables 3.2 and 3.3, pupils who did not take part in NAMER due to parental refusal are included in the category 'Absent'. In 2021, 0.2% of pupils did not participate in NAMER because their parents opted to withdraw them from the assessment. This proportion of pupils with parental refusal is similarly low to those observed in the previous two NAMER cycles.

Table 3.3. Response rates for the National Assessments, Sixth class Mathematics

| | 2009 | | 2014 | | 2021 | |
|----------------|----------|-------|----------|-------|----------|-------|
| | <i>N</i> | % | <i>N</i> | % | <i>N</i> | % |
| Exempt | 38 | 0.9 | 41 | 0.9 | 75 | 1.2 |
| Absent | 319 | 7.6 | 285 | 6.4 | 445 | 7.4 |
| Tested | 3832 | 91.5 | 4144 | 92.7 | 5516 | 91.4 |
| Total enrolled | 4189 | 100.0 | 4470 | 100.0 | 6036 | 100.0 |

The final sample used for the reporting of the NAMER 2021 results included only pupils who took both the test and the accompanying questionnaire. This decision was made to accommodate the scaling of the test data (for more information about the scaling see the section “Scaling of Test Data” presented in this chapter). There were 157 pupils in Second class and 121 pupils in Sixth class who took part in the assessment by answering reading or mathematics test items but did not complete the pupil questionnaire. This reduced the final sample size to 5044 pupils in Second class and 5395⁸ in Sixth class (about 89 % of the total number of enrolled pupils in the sampled schools).

Administration

Testing took place between May 4th and May 14th, 2021. The NAMER tests and questionnaires were administered by class teachers, while school coordinators were responsible for organising and overseeing assessments in each school. Usually, the school principal or another class teacher took on the role of school coordinator. A number of schools were also visited by an inspector from the Department of Education who quality assured the administration of the tests and reported back to the ERC. In previous cycles, inspectors would have delivered test materials to schools on the day of testing, and they would have overseen the test administration in all schools before returning materials to the ERC. However, due to the COVID-19 pandemic, inspectors visited only a proportion of schools (approximately 20 %) in 2021. Schools that received a visit were contacted by their assigned inspector prior to their test date to arrange and confirm testing dates.

Test booklets and other assessment materials were delivered to school coordinators in the week prior to the testing window. As NAMER is a secure test, used only for the National Assessments, school coordinators were asked to protect the security of all booklets and forms. Schools were asked to notify parents and guardians of the plan to administer the test using an information letter provided by the ERC in the week prior to testing. The letter informed parents and guardians of the purpose of the test and noted that parents could notify the school if they wished to withdraw their child from the test. Testing was conducted over one morning in each participating school. The total time required for the test (including breaks) was approximately 90 minutes, at both grade levels. Questionnaires were administered to pupils during the test window, often on the same day as the tests.

In Irish-medium schools, bilingual mathematics test booklets were administered at Sixth class. In such schools, the pupil, teacher and principal questionnaires were also bilingual. The decision as to whether the tests and questionnaires would be taken in Irish or English was made by individual pupils, teachers and school principals. The school principal and teacher questionnaires were also available in online format to accommodate the participants and improve response rates. Only a small number of teachers (two teachers at Sixth class) and one school principal completed the online version of the questionnaire.

⁸ The initial sample size was 5396, however, one case was excluded from the final sample at a later stage due to the fact that the pupil had not answered any of the pupil questionnaire items. The weighted sum of cases in 6th class is 5396.

Sampling weights

Sampling weights were calculated prior to the analysis of the test data. Weights are necessary since schools (and therefore pupils) were sampled disproportionately with regard to their overall presence in the population. The weighting process also applies a correction to account for non-response at school-level (e.g. a school not being able to participate) and at pupil-level (e.g., a pupil being absent on the day of testing). Weighting of data ensures that the contributions of certain groups of pupils (e.g. pupils attending large schools, or pupils who were present on the day of testing) are not over- or underrepresented in the data and therefore do not bias findings. Sampling weights feed into the scaling of test data and the analysis and reporting of data from the questionnaires.

The procedure for the calculation of weights was the same at both Second and Sixth class. The weights for NAMER 2021 were calculated as follows:

Equation 3.1.

$$W_{popwgt} = W_{sbw} * W_{scnr} * W_{cbw} * W_{pcnr}$$

Equation 3.2.

$$W_{sampwgt} = n * \frac{W_{sbw} * W_{scnr} * W_{cbw} * W_{pcnr}}{N}$$

Where:

- W_{popwgt} is the population weight (which sums to the size of the population)
- W_{sbw} is school base weight (the inverse of the probability of the school being selected)
- W_{scnr} is correction for non-response at the school level
- W_{cbw} is class base weight (the inverse of the probability of the class being selected)
- W_{pcnr} is correction for non-response at the pupil level.
- $W_{sampwgt}$ is the sample weight (the normalised version of the population weight which sums to the size of the sample)
- n is the number of pupils in the achieved sample
- N is number of pupils in the population

Sampling variance

NAMER, like many large-scale assessments, collects data from a sample of pupils instead of recruiting every individual in the target population. In any population, there are many possible samples that could be selected for a study, and each one of them would yield different results. It is important to account for this uncertainty due to sampling, which is commonly referred to as sampling variance. When cluster samples (where pupils are grouped within classes and schools) are selected, this uncertainty increases as the selected pupils may not represent the diversity of the population to the same extent as the same number of pupils from a simple random sample.

Sampling variance is estimated through the simulation of multiple samples based on the achieved sample. The larger the difference among the estimates of the same statistic that emerge from different samples, the larger the sampling variance. With simple random samples, the simulated distribution of statistics across multiple samples may be estimated based on the standard deviation from the current sample. However, with a complex cluster sample like that used by NAMER, the simulated distribution across

samples is estimated by replicating the estimation process across different combinations of pupils from the current sample.

There are different approaches to creating replication samples. One of the most straightforward and commonly used methods is the Jackknife approach (Efron, 1982; Wolter, 2007). Under this approach, sampled schools are grouped into zones, each containing two schools. For each replicate of the sample, one school is excluded from the sample (weighted to zero) while the weights of the remaining school in the same zone are doubled to make up for the omitted school, resulting in a number of new replicate samples⁹.

In 2021, the number of strata increased considerably; there were 23 strata, some of which had only one sampled school. Sampling zones were constructed within explicit strata. For strata with an even number of schools, the Jackknife zones were created as usual (schools were grouped in pairs). In strata with a single school or an odd number of schools, pupils in the unpaired schools were randomly divided to make up two “quasi” schools that create a new zone. Consequently, each sampling zone consisted of either a pair of schools or a single school whose pupils were split into two “quasi” schools.

In NAMER 2021, 167 schools participated in the assessment of Second class pupils resulting in 88 zones and 176 replicate weights. In Sixth Class, 168 schools took part in NAMER 2021 resulting in 89 zones and 178 replicate weights. As explained earlier, the replicate weights are used to simulate multiple subsamples based on the achieved sample selected for the purposes of the study. The variability among the estimates from these replicate samples is used to inform the standard errors of the full sample estimates.

The effect sizes produced by the data analyses are assumed to be drawn from a t-distribution with a mean equal to each estimate, variance equal to the sampling variance of the estimate, and degrees of freedom proportional to the number of replicates used to calculate the estimate. The properties of the t-distribution were used to estimate confidence intervals around estimates and determine critical values for testing the statistical significance of the difference between estimates.

This sampling variance estimation approach is also used by other large-scale assessments, such as TIMSS and PIRLS (Foy & LaRoche, 2020).

Scaling of test data

Similarly to previous NAMER cycles, the cognitive data (i.e., test data) were calibrated and scaled using the IRT approach. Compared with Classical Test Theory (CTT), IRT provides a more adaptable, effective, and efficient framework of test development and scaling that allows the unique characteristics of each test item to be taken into account. In CTT, the difficulty of items (and tests) depends on the ability of the sample of pupils taking an assessment, whereas in IRT, item characteristics are not dependent on the particular group of pupils taking an assessment (Hambleton et al., 1991).

Another advantage of IRT over CTT is its ability to effectively deal with incomplete testing designs. As mentioned earlier, for the purposes of NAMER, each pupil only sees a subset of the test items. The advantage of this approach is that a wider range of items can be used, thus improving the curriculum coverage and content validity without overburdening pupils with very long tests. However, as pupils see different sets of test items, steps need to be taken to ensure that pupil scores derived from different sets of test items are comparable to each other. Comparability of results from pupils taking different test booklets was ensured firstly by the random assignment of booklets (there should be no systematic differences between the ability levels of pupils taking any particular booklet) and secondly by the fact that all pupils within a grade level and domain were presented with a common set of items; for example, vocabulary items were common across all Second class reading booklets. The use of IRT also facilitates the process of dropping and adding items across cycles, while still allowing for the measurement of trends.

⁹ The number of the replicate samples is usually equal to the number of replicate weights which, in turn, is double the number of school zones.

Two-parameter (2PL) IRT models were used for the calibration and the scaling of the NAMER data using the Marginal Maximum Likelihood (MML) estimation approach. Pupils' ability (θ , or theta) was estimated by taking into account the responses of each pupil to each test item, as well as the difficulty (b) and discrimination (α) of each test item. The item difficulty and pupil ability estimates are on the same scale. Figures 3.1 and 3.2 show how IRT models the relationships between pupil ability, item difficulty, and item discrimination using item characteristic curves (ICCs) for two different items (item A and item B). Pupil ability is shown on the horizontal axis, and the probability of a correct answer on the vertical. The sloping line, rising smoothly from left to right, shows that for both items the probability of a correct response increases as the ability of pupils increases. Pupils at the low end of the ability scale have a low probability of answering the items correctly, while pupils at the high end of the ability scale are more likely to get the items right.

As indicated by the ICCs, item A is easier than item B, as pupils with the same ability level have higher chances of getting item A correct, compared to item B; for example, a pupil with a theta of 0 has about a 73 % chance of answering item A correctly but less than a 40 % chance of answering item B correctly. In general, item difficulty in IRT models is defined as the proficiency level at which a pupil has a 50 % chance of getting the correct answer. The ability level required to have a 50 % chance to get item A right ($b = -0.7$) is lower than that required for item B ($b = 0.7$).

The discrimination of an item refers to the ability of an item to differentiate between pupils whose ability is above the item difficulty versus those whose ability is below the item difficulty. A steep ICC slope indicates that the item discriminates well between these two groups of pupils, while a shallow slope indicates that it discriminates poorly. From the ICC graphs, we can see that item A has higher discrimination than item B. For example, a pupil with an ability level of 1 has a much higher probability of answering item A correctly ($p = .9$) compared to a pupil with a theta of -1 ($p = .4$). The probability difference for two pupils at these ability levels is smaller for Item B: $p = .55$ for a pupil with an ability of 1 versus $p = .2$ for a pupil with an ability of -1.

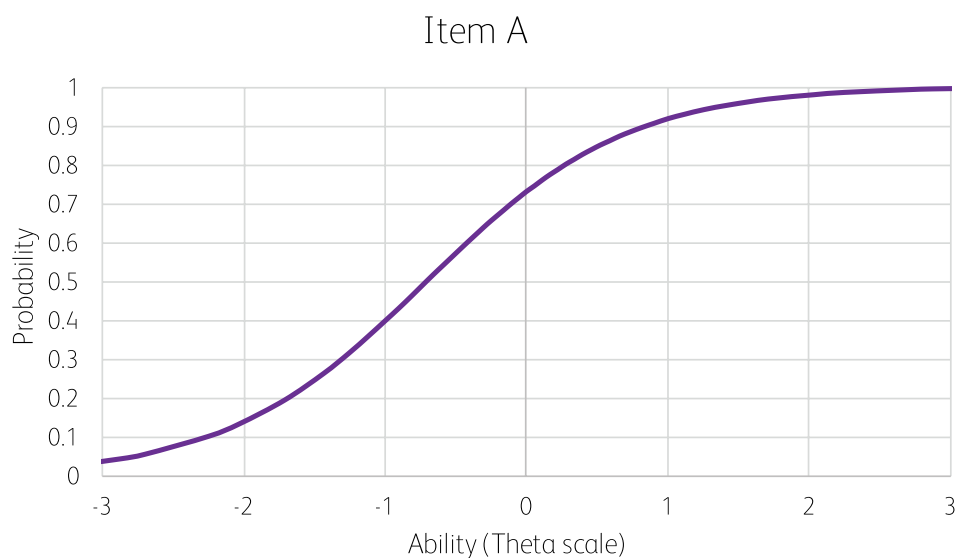


Figure 3.1. ICC, item A

Item B

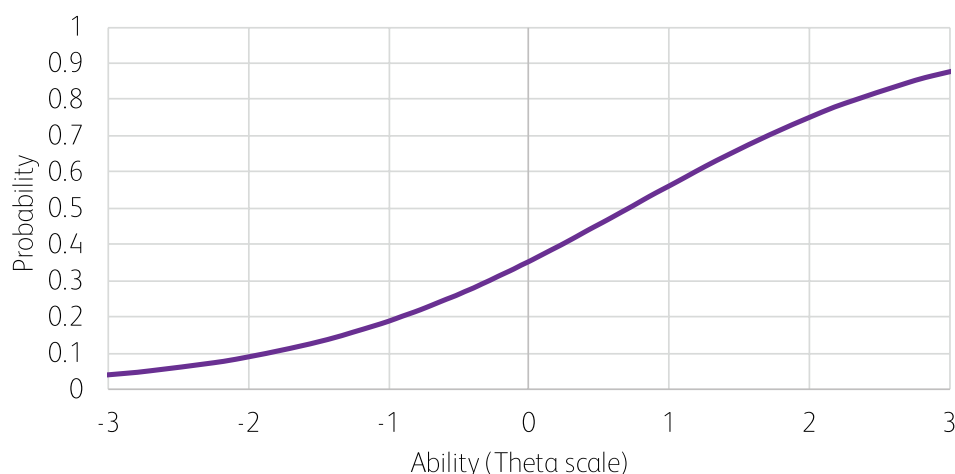


Figure 3.2. ICC, item B

All items, within each domain and level, were scaled on the basis of their difficulty and discrimination, and all pupil scores were calculated on the basis of these item parameters. Similar to previous cycles, concurrent scaling was used for NAMER 2021; that is, 2014 and 2021 item response data were scaled together. Overall, the IRT models fitted the data and the majority of the items performed well. In Second class reading, three trend items appeared to perform differently in 2021 compared to 2014; these were treated as new, instead of trend, items in 2021 as they were still performing well. One more item was excluded from the 2021 data because it did not perform as expected. In Sixth class Mathematics, one item was excluded from the analysis for the 2021 cycle because it performed significantly differently compared to the previous NAMER cycle.

In order to further improve the accuracy of the estimates of pupil performance and its relationship to contextual variables, some more advanced scaling techniques were implemented in NAMER 2021. More specifically, pupil scores were estimated using latent regression population models (commonly referred to as the conditioning model) that took into account all the available pupil-level contextual data (coming from the pupil questionnaire) as well as the school-level explicit stratification variable data¹⁰. Conditioning allows information about ability distributions of groups to increase the accuracy of ability estimation for individual pupils. Because pupils in different groups may have different ability distributions, the use of a latent regression model to estimate the conditioning model reduces underestimation of group differences by using different distributions of θ for pupils with different background characteristics or educational contexts. This process would be neither required nor wanted in the case of individual-level assessments where the focus is to get accurate performance estimates for individuals rather than for groups of pupils. However, since the aim of NAMER is to provide accurate group-level estimates, conditioning is useful for reporting unbiased results.

Additionally, in contrast to the previous NAMER cycle where each pupil was assigned a single performance estimate (score) for each domain (using the Expected A Priori [EAP] estimation method), in 2021 each pupil was assigned ten scores, called plausible values. Plausible values are random draws from the distribution of a pupil's estimated ability. Given the use of multiple test booklets containing different blocks of items and the relatively short length of the tests used in studies like NAMER, inferences about individual pupils' performance on the tests should be avoided. Instead, emphasis is put on the performance of the overall sample, as well as of certain subgroups. Plausible values allow us to represent the distributions of group performance more accurately than single scores.

¹⁰ The explicit stratification variables used in NAMER 2021 are: school size, DEIS status, area/language of instruction (Gaeltacht school, Gaelscoil, Ordinary School), type (junior, senior, vertical) and proportion of female pupils).

The use of conditioning variables in scaling and the generation of multiple plausible values allow for more accurate estimates of the means, standard deviations and standard errors of subgroups of interest, which in turn has consequences for the estimations of the statistical significance of the differences between groups.

The decision to use latent regression population models and generate plausible values to describe pupil performance in NAMER 2021 is in line with other national and international large-scale assessments (e.g., National Assessment of Educational Progress [NAEP], PISA, TIMSS and PIRLS).

Percent correct scores and plausible values were calculated for English reading and Mathematics. As well as overall test scores, scores were generated for each subdomain; reading components and processes, and mathematics content strands and process skills. Scores achieved by pupils participating in NAMER 2021 were projected onto the same scales as those used in 2014 and 2009 which allows the comparability of results across cycles.

The final calibration and scaling were conducted in a bespoke tool developed by Polymetrika International Inc. for the ERC. Checks were done with commercial software (e.g., IteMan [Assessment Systems Corporation, 2013] and Bilog-MG [Scientific Software International, 2003]) to verify the accuracy of the estimates generated through the Polymetrika International Inc. tool.

Data analysis

The IEA's International Database Analyzer (IDB Analyzer, version 5.0) was used for the estimation of means, percentages and other measures of central tendency and variability (IEA, 2022). IDB Analyzer is a free software tool which has been developed to analyse data from most major large-scale assessments (e.g., PISA, TIMSS and PIRLS). Analysis of the Irish National Assessment data is not directly supported by IDB Analyzer, so the NAMER data were formatted to be analysed as PIRLS and TIMSS data, for Second class English reading and Sixth class Mathematics respectively.

Appropriate sample weights, replicate weights and plausible values were used in the analysis to provide accurate estimates of the statistics and their standard errors. Tests of statistical significance and effect sizes, described earlier, were used to guide the interpretation of group differences. Correlation analysis was used to examine the relationship among variables with interval or ordinal scales.

Note on dependent and independent samples

Implementing the approach followed by other large-scale assessment studies (e.g., TIMSS, PIRLS and PISA), in NAMER we treated samples (i.e., groups) that derive from the same national sample, as well as their statistics, as dependent. For example, when examining the performance gap between girls and boys in the NAMER 2021 Sixth class mathematics test, the two gender groups are treated as dependent because they have been selected from the same national sample. Therefore, the standard errors of the difference between such groups were corrected to account for the fact that the two statistics are dependent.

It can be argued that groups deriving from explicit stratification variables (e.g., DEIS status) could be treated as independent as they are sampled from different sub-populations. However, the convention is to treat them as dependent, partly because the available analysis tools (e.g., IDB analyzer) cannot automatically identify the stratification variables to treat them differently. Both approaches (treating groups of explicit stratification variables within a country as either dependent or independent) are widely considered to be acceptable. In NAMER 2021, such groups were treated as dependent samples.

Comparisons of statistics deriving from different cycles of the study can be treated as independent. For example, when boys' performance in 2014 is compared to boys' performance in 2021, the SE of the difference between the two statistics can be estimated using the usual SED formula (see Equation 3.3), without applying the correction for dependent samples.

Equation 3.3.

$$SE_{(a-b)} = \sqrt{SE_a^2 + SE_b^2}$$

CHAPTER 4

Achievement Outcomes and Trends

This chapter describes the English reading achievement of Second class pupils and the mathematics achievement of Sixth class pupils who participated in NAMER 2021 and compares their performance with that of pupils who participated in the previous NAMER cycles. Throughout the chapter, outcomes are presented for NAMER 2009, 2014 and 2021. However, tests for statistical significance were only conducted between outcomes in NAMER 2021 and NAMER 2014. All effect sizes refer to the comparisons between these two NAMER cycles.

The chapter consists of two main sections. Second class reading performance is reported first, followed by Sixth class mathematics performance (overall, and by subscale). In each section, reading and mathematics performance are also described in terms of proficiency levels and considered in light of the updated performance targets set out in the Interim Review of the National Strategy to Improve Literacy and Numeracy (Department of Education and Skills, 2017b).

Second class English reading

Overall reading performance and performance by component

Pupils' overall English reading performance is examined, as well as performance on the two content areas (components): reading Vocabulary and reading Comprehension. As shown in Figure 4.1, in NAMER 2021, 63.2% of Vocabulary items and 62.3% of Comprehension items were answered correctly. Overall, pupils correctly answered 62.7% of the reading items that appeared on their tests.¹¹

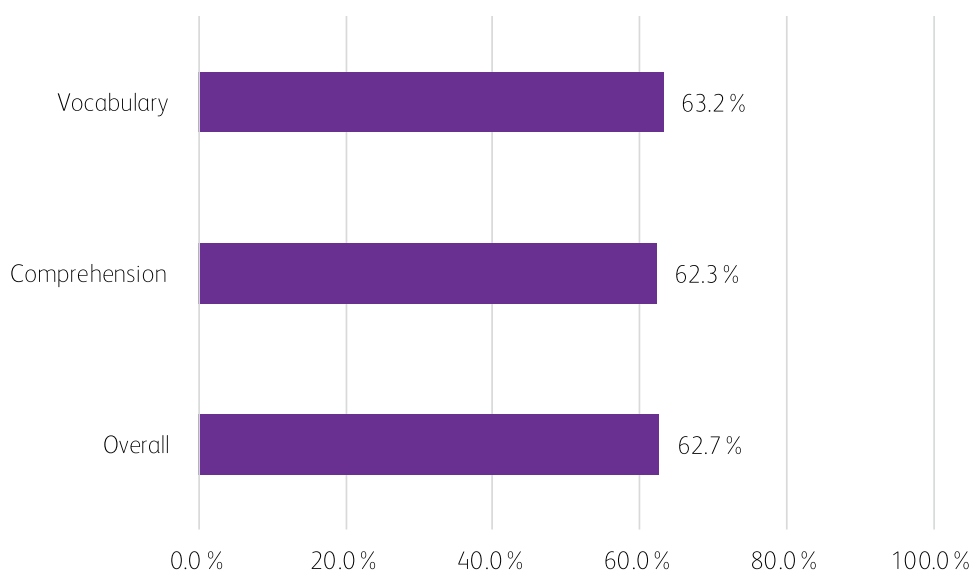


Figure 4.1. Mean percent correct scores by English Reading component, NAMER 2021

¹¹ As noted in Chapter 2 (Assessment Frameworks), a number of Vocabulary and Reading Comprehension items that had been used in NAMER 2014 were retired and replaced with new items in 2021. This is a common practice in large-scale standardised assessments that helps keep the test up-to-date and allows for items that do not perform as expected to be replaced with more robust items. The items that did not perform in 2014 were replaced with robust items with appropriate difficulty levels. Most of the new items in 2021 were more difficult than the retired ones from 2014.

Trend comparisons based solely on percent correct scores of the whole test may be misleading, because any cross-cycle changes in the difficulty of a test cannot be taken into account.¹² Scale scores generated using the IRT framework account for any changes in the difficulty of a test from one cycle to the next. Therefore, these unbiased scale scores are used to compare pupils' performance across assessment cycles.

Table 4.1 displays the mean scale scores for each reading component and cycle, along with the effect sizes (Cohen's *d*) of the performance gaps between 2014 and 2021. Pupils' overall reading performance in 2021 expressed as a scale score was 260.8, while the scale scores for Vocabulary and Comprehension were 262.6 and 259.2, respectively. Although these scores were lower than the scores in 2014, none of the differences are statistically significant. The relevant effect sizes are also small, suggesting that the results for 2021 are not meaningfully different from the results for 2014.

Table 4.1. Mean scale scores in English Reading by component and year

| | 2009 | 2014 | 2021 | <i>d</i> |
|----------------------|-------|-------|-------|----------|
| Vocabulary | 250.0 | 264.9 | 262.6 | 0.05 |
| Comprehension | 250.0 | 262.5 | 259.2 | 0.07 |
| Overall | 250.0 | 264.0 | 260.8 | 0.07 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Performance can also be described in terms of percentile markers, which represent points on a distribution of achievement. In Table 4.2, scores are shown for overall performance, Vocabulary performance and Comprehension performance, at the 10th, 25th, 50th, 75th and 90th percentiles, for three cycles of NAMER. The table shows the score required for a pupil to perform as well as or better than a certain proportion of their peers. For example, pupils who performed at the 90th percentile performed as well as or better than 90 % of their peers. In 2021, pupils performing at the 90th percentile had an overall Reading score of 324.4. There are no statistically significant differences in scores at key percentile points between 2014 and 2021.

Table 4.2. Scores at key percentile points on the overall reading, vocabulary and comprehension scales by year

| | | Percentile rank | | | | |
|----------------------|-------------|-----------------|-------|-------|-------|-------|
| | | 10th | 25th | 50th | 75th | 90th |
| Vocabulary | 2009 | 186.6 | 208.9 | 247.3 | 284.7 | 319.1 |
| | 2014 | 197.4 | 229.0 | 264.6 | 299.4 | 331.7 |
| | 2021 | 199.2 | 228.0 | 262.3 | 297.1 | 326.7 |
| Comprehension | 2009 | 185.0 | 210.6 | 247.2 | 285.9 | 317.1 |
| | 2014 | 199.7 | 228.9 | 262.0 | 295.7 | 323.4 |
| | 2021 | 197.0 | 224.5 | 260.0 | 292.2 | 321.0 |
| Overall | 2009 | 186.4 | 211.4 | 246.6 | 282.4 | 318.8 |
| | 2014 | 199.9 | 229.7 | 261.6 | 296.7 | 328.1 |
| | 2021 | 197.3 | 225.3 | 260.9 | 294.1 | 324.4 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

¹² However, performance on trend items (those that were used across two or more cycles) can provide more accurate information and can facilitate percent correct score comparisons across cycles. Table A4.2 and A4.3 in the Appendix for Chapter 4 provide mean difficulty of the trend items across all three NAMER cycles and across the last two cycles, respectively.

Reading performance by process skill

Reading performance can also be categorised according to the three process skills examined in the assessment – Retrieve, Infer, and Interpret & Integrate.¹³ In NAMER 2021, 65.4 % of items on the Retrieve subscale were answered correctly, compared with 62.1 % and 56.8 % of items on the Infer and Interpret & Integrate subscales, respectively (Figure 4.2).

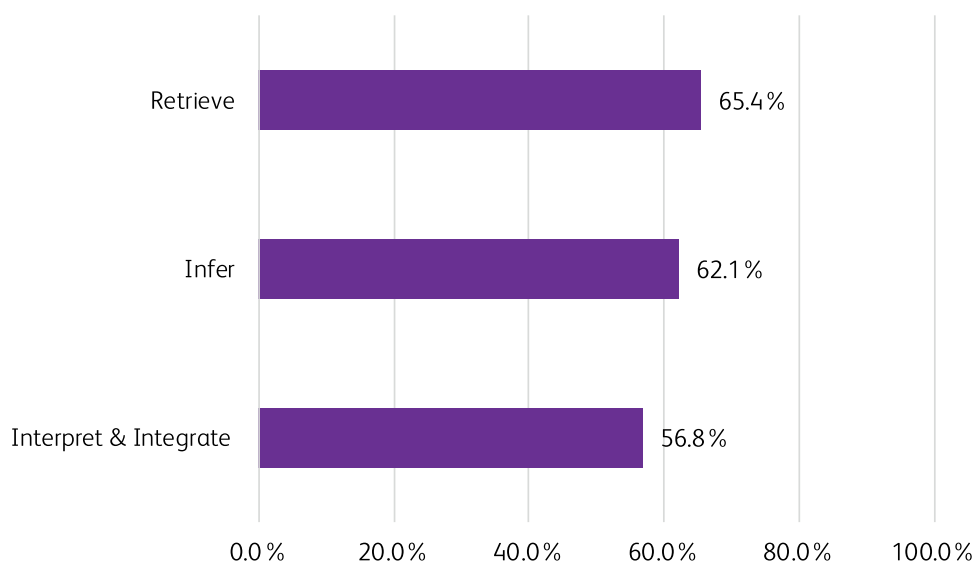


Figure 4.2. Mean percent correct score by reading process skill, NAMER 2021

Table 4.3 displays the mean scale scores for the process skills across the three cycles of NAMER. Mean scale scores for all three process skills were slightly lower in 2021 than in 2014; however, similarly to the content domains, none of the differences in mean scores are statistically significant and the relevant effect sizes are small, suggesting pupils in 2021 performed at a similar level to their counterparts in 2014.

Table 4.3. Mean scale scores by reading process skill and year, Second class

| | 2009 | 2014 | 2021 | <i>d</i> |
|----------------------------------|-------|-------|-------|----------|
| Retrieve | 250.0 | 260.5 | 257.8 | 0.06 |
| Infer | 250.0 | 263.2 | 260.9 | 0.05 |
| Interpret & Integrate | 250.0 | 261.1 | 260.1 | 0.02 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

In Table 4.4, scores at key percentile points are displayed for the Retrieve, Infer and Interpret & Integrate subscales, from NAMER 2009, 2014, and 2021. Pupils' scores at the 75th percentile on the Retrieve and Interpret & Integrate subscales are statistically significantly lower than the corresponding 2014 scores. Conversely, the score of the highest achieving pupils (at the 90th percentile) on the Interpret & Integrate subscale in 2021, is statistically significantly higher than the corresponding score in 2014.

¹³ Only two test items on the Second class Reading tests assessed the fourth process skill, Examine & Evaluate. Therefore, a subscale for this process skill was not created.

Table 4.4. English reading scores at key percentile points by process skill and year

| | | Percentile rank | | | | |
|-----------------------|------|-----------------|-------|-------|--------------|--------------|
| | | 10th | 25th | 50th | 75th | 90th |
| Retrieve | 2009 | 182.8 | 209.8 | 248.0 | 288.7 | 321.2 |
| | 2014 | 197.8 | 227.1 | 260.8 | 296.4 | 320.4 |
| | 2021 | 196.5 | 225.4 | 258.5 | 290.0 | 317.4 |
| Infer | 2009 | 180.2 | 212.3 | 250.0 | 288.5 | 319.6 |
| | 2014 | 198.4 | 228.6 | 264.9 | 299.6 | 322.7 |
| | 2021 | 198.5 | 226.8 | 261.4 | 294.3 | 322.0 |
| Interpret & Integrate | 2009 | 174.8 | 207.5 | 251.7 | 295.3 | 313.5 |
| | 2014 | 201.0 | 231.8 | 260.8 | 296.6 | 313.0 |
| | 2021 | 198.9 | 227.7 | 260.9 | 292.0 | 320.7 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Proficiency levels

Pupil performance can also be expressed in terms of proficiency levels, which represent clusters of skills and describe the types of tasks which pupils at different levels of performance can consistently complete. Pupils performing at Level 1 would be expected to successfully complete only the most basic tasks expected of their grade level, while those performing at Level 4 would be expected to successfully complete the most complex tasks. Pupils who do not reach Level 1 are not consistently able to display the skills assessed by the simplest items on the test.

In the following sections, Second class pupils' performance on the 2021 English reading test is reported with reference to the proficiency levels that were developed for the 2009 National Assessments and used for reporting on the 2014 National Assessments. Table 4.5 outlines the reading skills that Second class pupils at each proficiency level are expected to demonstrate, and the percentages of pupils performing at each proficiency level in NAMER 2009, 2014 and 2021.

The percentage of pupils performing at Level 4 was lower in NAMER 2021 than in 2014, and the percentage of pupils performing at or below Level 1 in NAMER 2021 was higher than in 2014. However, these differences are not statistically significant, meaning that the percentages of high- and low-performing pupils in both cycles are similar.

Table 4.5. Percentages of Second class pupils at each proficiency level on the overall English Reading scale by year

| Level | Score range | What pupils can typically do | Percentage of pupils | | |
|--------------|-------------|---|----------------------|------|-------------|
| | | | 2009 | 2014 | 2021 |
| 4 | 320+ | As well as succeeding on lower proficiency level skills, pupils at Level 4 can retrieve complex information (e.g., the information needed is located in multiple parts of the text). They can link multiple pieces of information to draw inferences. They can integrate text-wide information in order to identify the main themes in a text. As well as using discrete or explicit information, they can use the text as a whole to interpret character behaviour. | 10.0 | 13.5 | 11.7 |
| 3 | 269-319 | As well as Level 1 and 2 skills, pupils can process texts at a whole-text level, in order to retrieve information. They can make basic-level inferences, sometimes linking one or two discrete pieces of information. They can infer word meanings if the context provides clear clues. | 25.0 | 32.0 | 32.4 |
| 2 | 225-268 | As well as Level 1 skills, pupils can retrieve explicitly stated information where the wording of the question and the text differ. However, the information sought must be specific to a small section of text. They can make low-level inferences, including character motives, if the required information is explicitly stated in a specific section of the text. | 30.0 | 32.8 | 31.4 |
| 1 | 187-224 | Level 1 pupils show basic Reading skills. They can retrieve simple, explicitly stated, pieces of information, when there is a direct match between the wording of the question and the text. They are most successful on tasks that require comprehension of smaller units of text, such as sentences. They can perform some very basic interpretation and integration of text (e.g., identifying the theme of a text, where the theme is explicitly stated in the text). | 25.0 | 16.5 | 18.4 |
| <1 | <187 | Pupils below proficiency Level 1 have a less than 62.5 % chance of correctly answering a Level 1 question. Their Reading skills are very low, relative to other 2nd class pupils and are not properly assessed by the National Assessment. | 10.0 | 5.1 | 6.0 |

Note. 2021 percentages in bold are statistically significantly different from the corresponding 2014 percentage.

In 2011, performance targets related to proficiency levels were set out in the National Strategy for Literacy and Numeracy (Department of Education and Skills, 2011). Pupils' performance surpassed these targets in NAMER 2014. Following an interim review of the Strategy (Department of Education and Skills, 2017b), new targets were set for the next cycle of the National Assessments. This is demonstrated in Figure 4.3, which displays the original targets set in the 2011 Strategy and the results of NAMER 2014. Figure 4.3 also shows the revised targets set in 2017 and the results from NAMER 2021.

At the lower end of the achievement distribution, the target was to reduce the percentage of pupils performing at or below Level 1 to 20%. This target was not reached in 2021; the percentage of pupils performing at this level was statistically significantly higher (24.4%) than the target. At the upper end of the achievement distribution, 44.1% of pupils performed at or above proficiency Level 3 in 2021. This is a statistically significantly lower proportion of pupils than the target of 50%. In both cases, the new targets set in the reviewed National Strategy were not reached (Table A4.8).

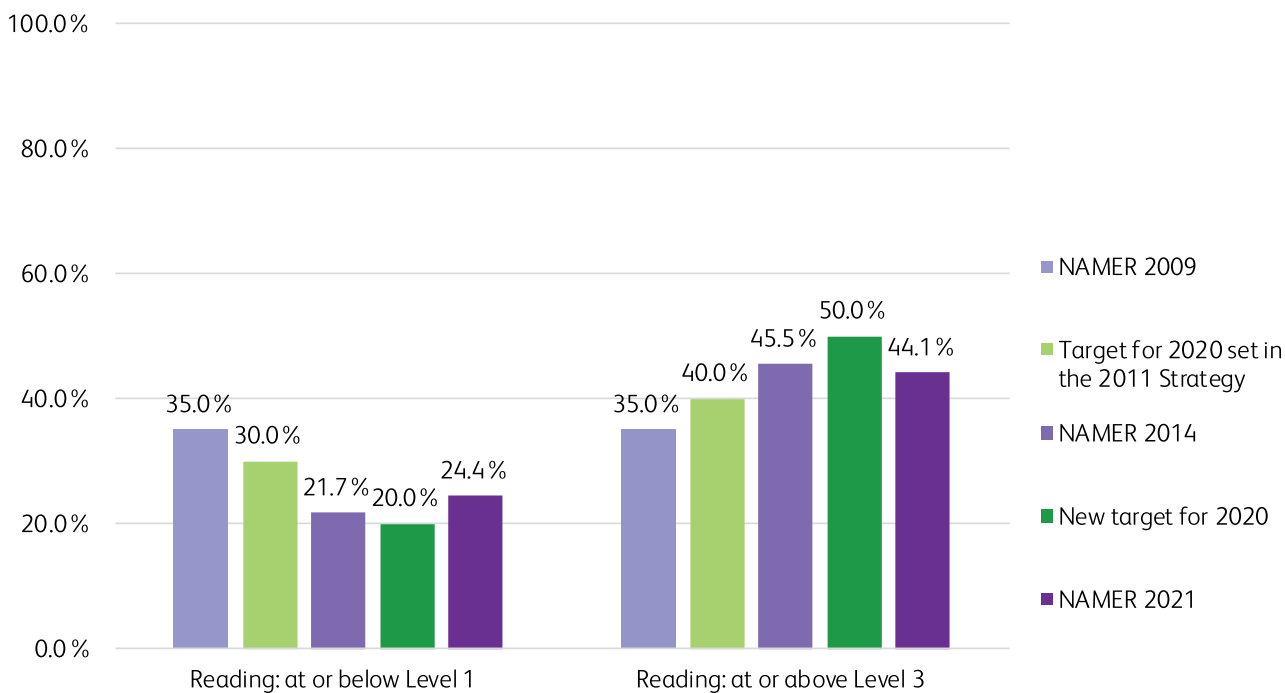


Figure 4.3. Percentages of Second class pupils achieving at various proficiency levels across years with old and new targets

Sixth class Mathematics

Overall mathematics performance and performance by content area

In this section, pupils' overall mathematics performance is examined, along with performance on the four content areas: Number & Algebra, Shape & Space, Measures and Data. As shown in Figure 4.4, overall, pupils answered 59.4% of items on the mathematics assessment correctly. Percent correct scores on content subscales ranged from 44.5% on Measures, to 60.8% on Shape & Space, 61.4% on Data and 63.1% on Number & Algebra.

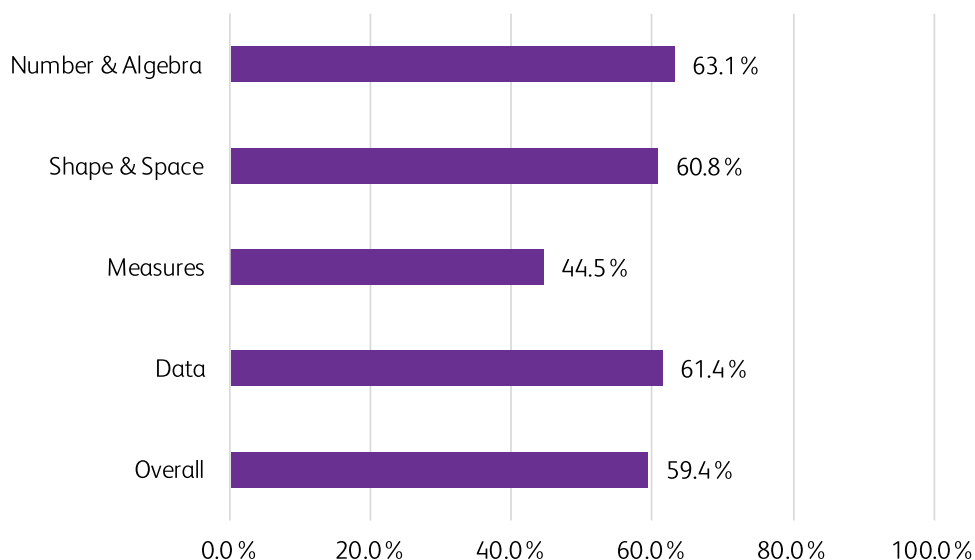


Figure 4.4. Mean percent correct scores by mathematics content area, NAMER 2021

As mentioned in the English reading section of this chapter, trend comparisons based solely on percent correct scores of the whole test may be misleading, because any cross-cycle changes in the difficulty of a test cannot be taken into account.¹⁴ Therefore, scale scores are used to compare pupils' performance across assessment cycles.

Table 4.6 displays the mean scale scores for each mathematics content area and cycle, along with the effect sizes (Cohen's *d*) of the performance gaps between 2014 and 2021. Pupils' overall mathematics performance in 2021 expressed as a scale score was 260.5, while the scale scores for the content areas were 259 for Number & Algebra, 256.4 for Shape & Space, 257.7 for Measures and 257.3 for Data. Although all scale scores were lower than the scores in 2014, only the scale score for Shape & Space is statistically significantly lower than the result in 2014. Overall, and for all subscales, effect sizes are small, suggesting that the results are not meaningfully different from results in 2014.

Table 4.6. Mean scale scores in Mathematics by content area and year

| | 2009 | 2014 | 2021 | <i>d</i> |
|-----------------------------|-------|-------|--------------|----------|
| Number & Algebra | 250.0 | 260.8 | 259.0 | 0.04 |
| Shape & Space | 250.0 | 262.8 | 256.4 | 0.14 |
| Measures | 250.0 | 258.5 | 257.7 | 0.01 |
| Data | 250.0 | 259.0 | 257.3 | 0.03 |
| Overall | 250.0 | 261.7 | 260.5 | 0.02 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

As has been explained earlier in this chapter, performance can also be described in terms of percentile markers, which represent points on a distribution of achievement. In Table 4.7, scores are shown for overall performance and content performance at the 10th, 25th, 50th, 75th and 90th percentiles, for three cycles of NAMER. The table shows the score required for a pupil to perform as well as, or better than, a certain

¹⁴ Performance on trend items (those that were used across two or more cycles) can provide more accurate information and can facilitate percent correct score comparisons across cycles. Tables A4.10 and A4.11 in the Appendix provide mean difficulty of the trend items across all three NAMER cycles and across the last two cycles, respectively.

proportion of their peers. For example, pupils who performed at the 90th percentile performed as well as or better than 90 % of their peers.

In 2021, pupils performing at the 90th percentile had an overall maths scale score of 325. Overall, there are no statistically significant differences in scores at key percentile points between 2014 and 2021. Only the Shape & Space subscale showed statistically significant differences in scale scores between 2014 and 2021; the scale scores for pupils at the 50th percentile (256.4) and 75th percentile (288.4) are statistically significantly lower than the corresponding scale scores in 2014.

Table 4.7. Mathematics achievement for content areas at key percentile points, by year, Sixth class

| | | Percentile rank | | | | |
|------------------|------|-----------------|-------|--------------|--------------|-------|
| | | 10th | 25th | 50th | 75th | 90th |
| Number & Algebra | 2009 | 184.9 | 212.8 | 247.7 | 286.8 | 315.2 |
| | 2014 | 195.9 | 225.7 | 260.2 | 295.6 | 324.5 |
| | 2021 | 195.0 | 224.6 | 259.4 | 293.4 | 322.4 |
| Shape & Space | 2009 | 183.6 | 211.9 | 250.3 | 288.0 | 315.6 |
| | 2014 | 197.3 | 230.9 | 264.0 | 297.4 | 322.0 |
| | 2021 | 197.9 | 224.9 | 256.4 | 288.4 | 315.2 |
| Measures | 2009 | 184.8 | 210.4 | 245.9 | 286.3 | 319.6 |
| | 2014 | 191.9 | 218.0 | 254.9 | 295.9 | 326.7 |
| | 2021 | 192.5 | 222.4 | 257.6 | 293.7 | 323.9 |
| Data | 2009 | 180.7 | 212.9 | 251.7 | 290.0 | 311.6 |
| | 2014 | 192.3 | 226.3 | 261.7 | 296.1 | 325.2 |
| | 2021 | 193.4 | 222.7 | 257.6 | 291.9 | 321.9 |
| Overall | 2009 | 183.4 | 213.8 | 248.7 | 285.6 | 314.5 |
| | 2014 | 198.0 | 227.3 | 261.3 | 296.6 | 326.0 |
| | 2021 | 195.5 | 225.5 | 260.9 | 295.5 | 325.0 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Mathematics performance by process skill

Mathematics performance can also be categorised according to the process used to answer the question. The mathematics processes examined in this assessment were Understand & Recall, Implement, Reason, Integrate & Connect and Apply & Problem Solve. As shown in Figure 4.5, items assessing Apply & Problem Solve were the most difficult, with a mean percent correct score of 53.2 %. Integrate & Connect items were the easiest, with pupils answering 66.3 % of these items correctly.

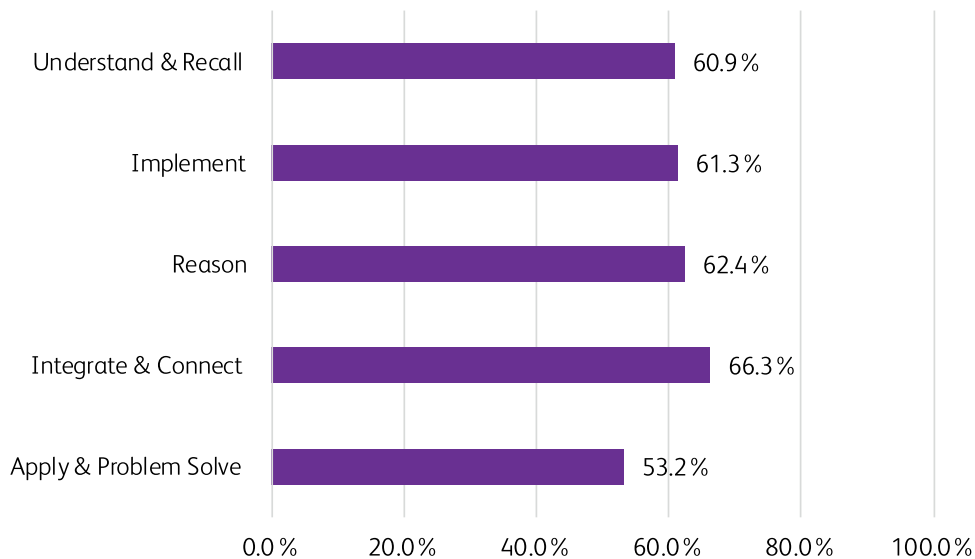


Figure 4.5. Mean percent correct scores by mathematics process skill, NAMER 2021

Table 4.8 shows the scale scores for the process skills across the three cycles of NAMER. Compared with 2014, scores were slightly lower on all process subscales in 2021. However, none of the differences are statistically significant and the corresponding effect sizes are small. This indicates that performance was similar across the two cycles.

Table 4.8. Mean scale scores in Mathematics by process skill and year, Sixth class

| | 2009 | 2014 | 2021 | <i>d</i> |
|----------------------------------|-------|-------|--------------|----------|
| Understand & Recall | 250.0 | 260.2 | 258.0 | 0.05 |
| Implement | 250.0 | 262.8 | 259.9 | 0.06 |
| Reason | 250.0 | 261.6 | 260.4 | 0.02 |
| Integrate & Connect | 250.0 | 257.4 | 257.3 | 0.00 |
| Apply & Problem Solve | 250.0 | 259.2 | 258.3 | 0.02 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Performance can also be described in terms of percentile markers, as was shown earlier in this chapter. Table 4.9 shows the mean scores at the 10th, 25th, 50th, 75th and 90th percentiles for each process skill in 2009, 2014 and 2021. There are no statistically significant differences in the scores achieved at key percentile points between 2014 and 2021, with the exception of the Integrate & Connect subscale. In 2021, the mean score at the 75th percentile is significantly lower than the corresponding score in 2014, while at the 25th and 90th percentiles, it is significantly higher. There is no clear pattern of change in performance, as at one percentile point, the score is statistically significantly lower in 2021 while at two others, it is higher.

Table 4.9. Mathematics achievement for process skills at key percentile points, by year, Sixth class

| | | Percentile rank | | | | |
|-----------------------|------|-----------------|--------------|-------|--------------|--------------|
| | | 10th | 25th | 50th | 75th | 90th |
| Understand & Recall | 2009 | 175.9 | 216.8 | 257.3 | 297.3 | 306.0 |
| | 2014 | 195.1 | 227.6 | 262.4 | 293.9 | 321.3 |
| | 2021 | 196.3 | 225.9 | 258.8 | 290.9 | 318.2 |
| Implement | 2009 | 180.5 | 215.2 | 251.8 | 286.6 | 312.2 |
| | 2014 | 196.2 | 226.9 | 264.8 | 300.1 | 326.6 |
| | 2021 | 195.0 | 224.9 | 261.3 | 294.7 | 323.4 |
| Reason | 2009 | 184.8 | 213.1 | 248.3 | 284.7 | 319.2 |
| | 2014 | 195.0 | 227.9 | 263.1 | 296.9 | 325.0 |
| | 2021 | 193.6 | 225.9 | 261.1 | 296.1 | 325.4 |
| Integrate & Connect | 2009 | 181.8 | 217.0 | 250.6 | 285.8 | 316.3 |
| | 2014 | 196.6 | 216.2 | 261.9 | 298.7 | 304.1 |
| | 2021 | 193.4 | 224.5 | 257.7 | 290.4 | 319.2 |
| Apply & Problem Solve | 2009 | 182.4 | 211.0 | 249.5 | 284.8 | 317.5 |
| | 2014 | 192.8 | 222.6 | 257.3 | 294.4 | 325.4 |
| | 2021 | 193.1 | 223.5 | 258.3 | 293.3 | 322.3 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Proficiency Levels

Table 4.10 outlines the mathematics skills that Sixth class pupils can demonstrate at each proficiency level, and the percentages of pupils performing at each level in NAMER 2009, 2014 and 2021. The percentages of pupils performing at each proficiency level have remained relatively stable between 2014 and 2021. No statistically significant differences were observed in the percentages of pupils achieving at each proficiency level between 2014 and 2021.

Table 4.10. Percentages of pupils at each proficiency level on the overall mathematics scale by year

| Level | Score range | What pupils can typically do | Percentage of pupils | | |
|-------|-------------|--|----------------------|------|------|
| | | | 2009 | 2014 | 2021 |
| 4 | 316+ | Pupils at Level 4 can multiply and divide decimals by decimals, and carry out simple algebraic procedures involving evaluation of linear expressions and one-step equations. They can demonstrate a high level of understanding of signed integers and number theory concepts such as prime and composite numbers. They can deduce symbolic rules for simple functions. At this level pupils can also analyse geometric shapes in detail and deduce rules about them. They can construct circles. They can plot coordinates and use scales on maps or plans to calculate distances and areas. They can solve non-routine and multi-step practical problems involving ratios, mixed numbers, percentage gain or loss, value for money comparisons, currency conversions, speed, and time zones. | 10.0 | 14.9 | 14.1 |
| 3 | 273-315 | Pupils at Level 3 can add and subtract mixed numbers and decimals. They can demonstrate understanding of decimal notation, factors and multiples, exponents, and square roots. They can connect verbal and symbolic representations of word problems. They can construct and measure angles and construct triangles and rectangles given selected sides and angles. Pupils at this level can classify triangles and quadrilaterals based on angle and line properties and rules. They can identify properties of 3-D shapes. They can manipulate commonly used units of area, capacity and weight. They can read, interpret, and analyse pie-charts, multiple-bar bar-charts and trend graphs. They can estimate simple probabilities. They can solve routine and non-routine word problems involving operations with fractions, decimals and percentages, length and perimeter, capacity, and time. | 25.0 | 27.2 | 27.3 |
| 2 | 230-272 | Pupils at Level 2 can multiply fractions and decimals, estimate products, calculate common factors and multiples of whole numbers, and convert fractions and decimals to percentages. They can identify prime numbers within 30 and identify rules for number patterns. They can demonstrate understanding of a letter as a placeholder in algebraic expressions, and complete two-step number sentences involving addition and subtraction. Pupils at this level can construct lines and circles, estimate angles and use properties of shapes to calculate line and angle sizes. They can make logical deductions from simple data sets. They can solve multi-step word problems involving operations with integers, fractions and percentages. | 30.0 | 31.4 | 31.3 |
| 1 | 184-229 | Pupils at Level 1 can add, subtract, and round whole numbers and decimals. They show understanding of whole number notation and can connect numeric and verbal representations of large numbers. Pupils at this level can classify angles and identify templates of simple 3-D shapes. They can manipulate commonly used units of length. They can read and interpret, without calculation, simple frequency tables, pie-charts, bar charts and trend graphs. They can solve routine word problems involving the four operations with whole numbers. | 25.0 | 20.9 | 21.1 |
| <1 | <184 | Pupils below proficiency Level 1 have a less than 62.5% chance of correctly answering a Level 1 question. Their mathematical skills are very low, relative to other 6th class pupils and are not properly assessed by the National Assessments. | 10.0 | 5.6 | 6.1 |

Figure 4.6 displays the original mathematics performance targets set in the 2011 Strategy and the results of NAMER 2014, along with the revised targets set in 2017 and the results from NAMER 2021.

At the lower end of the achievement distribution, the target was to reduce the percentage of pupils performing at or below Level 1 to 20%. This target was not reached in 2021; the percentage of pupils performing at this level was statistically significantly higher (27.3%) than the target. At the upper end of the achievement distribution, 41.4% of pupils performed at or above proficiency Level 3 in 2021. This is a statistically significantly lower proportion of pupils than the target of 50%. In both cases, the new targets set in the revised National Strategy were not reached (Table A4.16).

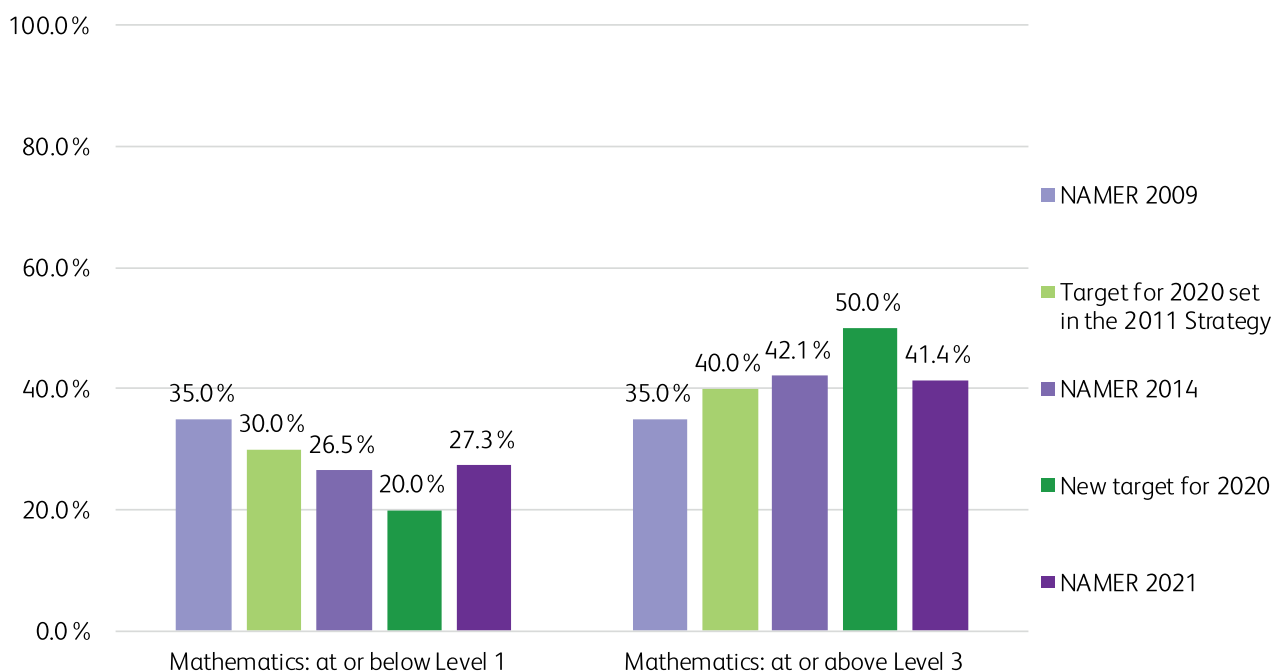


Figure 4.6. Percentages of Sixth class pupils achieving at various proficiency levels across years with old and new targets

Summary

English reading

Overall mean scores for Second class reading in NAMER 2021 were not statistically significantly different from the corresponding scores in NAMER 2014. Likewise, no significant changes in performance were observed on either of the subcomponents (Vocabulary or Comprehension) or any of the process skill subscales (Retrieve, Infer, Interpret & Integrate).

There are no statistically significant differences in overall reading scores at key percentile points between 2014 and 2021, but some subscales (Retrieve, Interpret & Integrate) show statistically significant differences.

The percentages of pupils performing at the different proficiency levels did not change statistically significantly in 2021, compared to 2014. Therefore, the performance targets for Second Class reading set out in the Interim Review of the National Strategy to Improve Literacy and Numeracy, based on the proportions of pupils performing at the highest and lowest proficiency levels, were not met.

Mathematics

Sixth class pupils' overall performance on Mathematics in NAMER 2021 was lower than in 2014. Performance on all content area and process skill subscales was also lower in 2021 compared with 2014. However, none of these differences is statistically significant apart from mean performance on the Shape & Space content subscale.

There are no statistically significant differences in overall scores at key percentile points between 2014 and 2021. However, there are a few significant differences in some subscales. Also, the percentages of pupils performing at the different proficiency levels did not change statistically significantly in 2021, compared with 2014. The performance targets for Sixth Class Mathematics set out in the Interim Review of the National Strategy to Improve Literacy and Numeracy (Department of Education & Skills, 2017), based on the proportions of pupils performing at the highest and lowest proficiency levels, were not reached. However, no statistically significant changes in performance are observed between 2014 and 2021 at each proficiency level overall.

CHAPTER 5

Performance and Gender

This chapter examines performance in English reading and Mathematics by gender and year. First, findings related to Second class reading are presented. Then, findings related to Sixth class Mathematics are presented. Findings for each subject are presented in two sections – firstly, results from the current cycle, NAMER 2021, are presented and comparisons between gender groups are examined. Secondly, trend comparisons between 2014 and 2021 are presented.

It should be noted that in previous NAMER cycles, pupils were only grouped by two genders, Boy and Girl. However, in 2021, three options were given to pupils – Boy, Girl and Other. Table 5.1 shows the number and percentage of pupils in each gender category in the NAMER 2021 sample. Results for pupils who selected ‘Other’ are displayed in the tables, but tests for statistically significant differences between this and the other gender groups were not conducted due to the very small number of pupils who selected this option. Statistical significance tests were only conducted between boys and girls.

Table 5.1. Number of pupils in each gender category, NAMER 2021

| | Second class | | Sixth class | |
|--------------|--------------|------|-------------|------|
| | <i>N</i> | % | <i>N</i> | % |
| Boy | 2560 | 50.9 | 2618 | 48.7 |
| Girl | 2452 | 48.8 | 2697 | 50.1 |
| Other | 16 | 0.3 | 65 | 1.2 |

Second class English reading performance

This section presents Second class pupils’ English reading performance in NAMER 2021, and compares boys’ and girls’ performance. First, pupil achievement is considered in terms of percent correct and scale scores on the overall reading scale, as well as by content area and process skill. Then, achievement is outlined with reference to proficiency levels.

Performance on overall reading scale, by component and process skill

On the overall reading scale in 2021, 60.8 % of test items were answered correctly by boys. The corresponding figure for girls was 64.6 %. Girls also outperformed boys on the Vocabulary and Comprehension sections of the reading test.

Boys correctly answered a slightly higher proportion of Vocabulary items (62.3 %) than Comprehension items (59.9 %). The proportions of items answered correctly by girls on the Vocabulary and Comprehension subscales were 64.1 % and 64.9 %, respectively. Pupils who selected ‘Other’ as their gender correctly answered 62.5 % of items on the overall scale, 62.1 % on the Vocabulary subscale and 62.9 % on the Comprehension subscale.

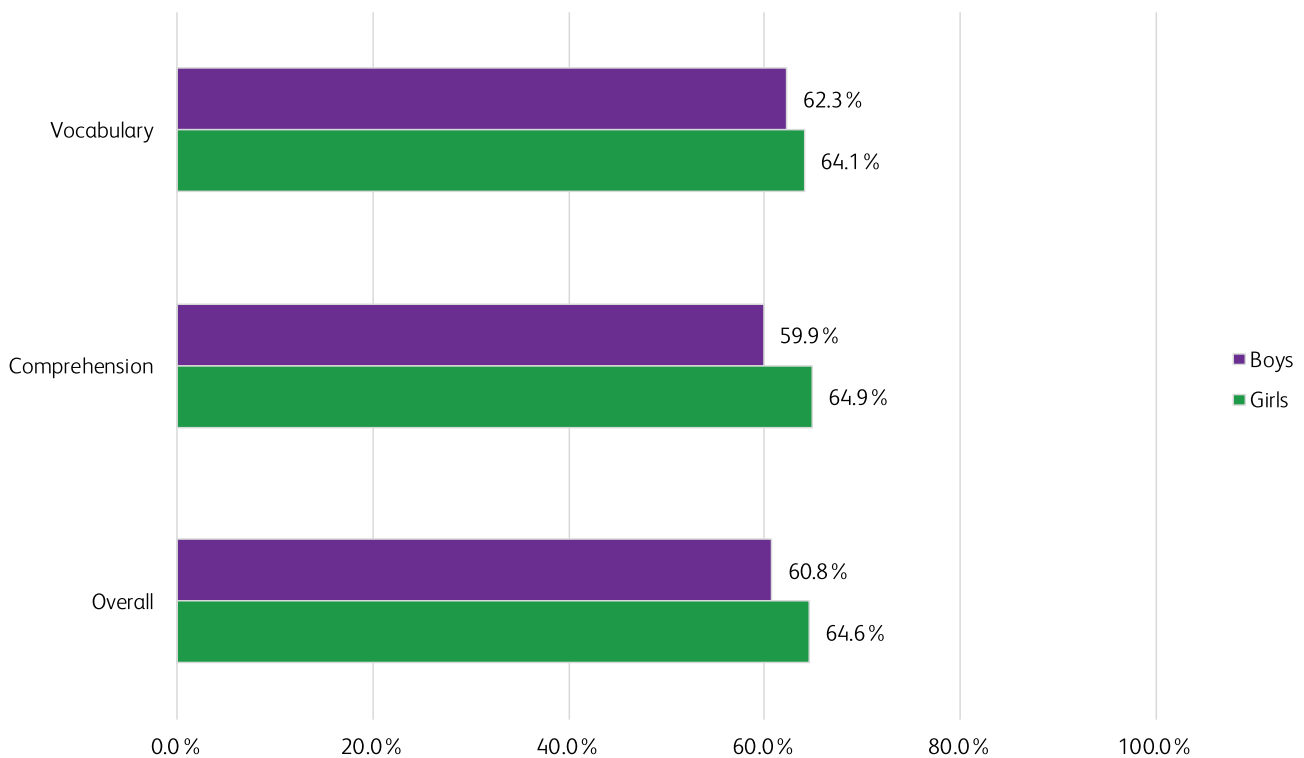


Figure 5.1. Mean percent correct scores by English Reading component and gender, NAMER 2021

The scale scores show that in 2021, girls outperformed boys on the overall reading scale by 8.3 score points, which is a statistically significant difference. There is no statistically significant difference in performance between girls and boys on the Vocabulary subscale. However, as shown in table 5.2, girls achieved a mean score on the Comprehension subscale that is statistically significantly higher than boys by a difference of 9.1 score points. Pupils who selected ‘Other’ as their gender achieved a mean score of 261.1 on the overall scale, and their performance on the Vocabulary subscale (261.2) was similar to their performance on Comprehension subscale (260.4). The effect sizes of the performance gaps between boys and girls are small.

Table 5.2. Mean scale scores in English Reading by component and gender, NAMER 2021

| | Boys | Girls | Boys - Girls |
|----------------------|-------------------------|--------------|--------------|
| | <i>Mean scale score</i> | | <i>d</i> |
| Vocabulary | 260.8 | 264.6 | 0.08 |
| Comprehension | 254.8 | 263.9 | 0.19 |
| Overall | 256.8 | 265.1 | 0.17 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Girls’ scores in bold are statistically significantly different from the corresponding boys’ score.

Figure 5.2 displays the proportion of items answered correctly by each gender group on each process skill subscale in 2021. Items on the Interpret & Integrate subscale appeared to be the most difficult across all gender groups, and items on the Retrieve subscale appeared to be the easiest for boys and girls. The subscale with the highest percentage of items answered correctly by pupils who selected ‘Other’ as their gender was Infer (64.3%). Girls correctly answered a higher proportion of items than boys on all three process skill subscales.

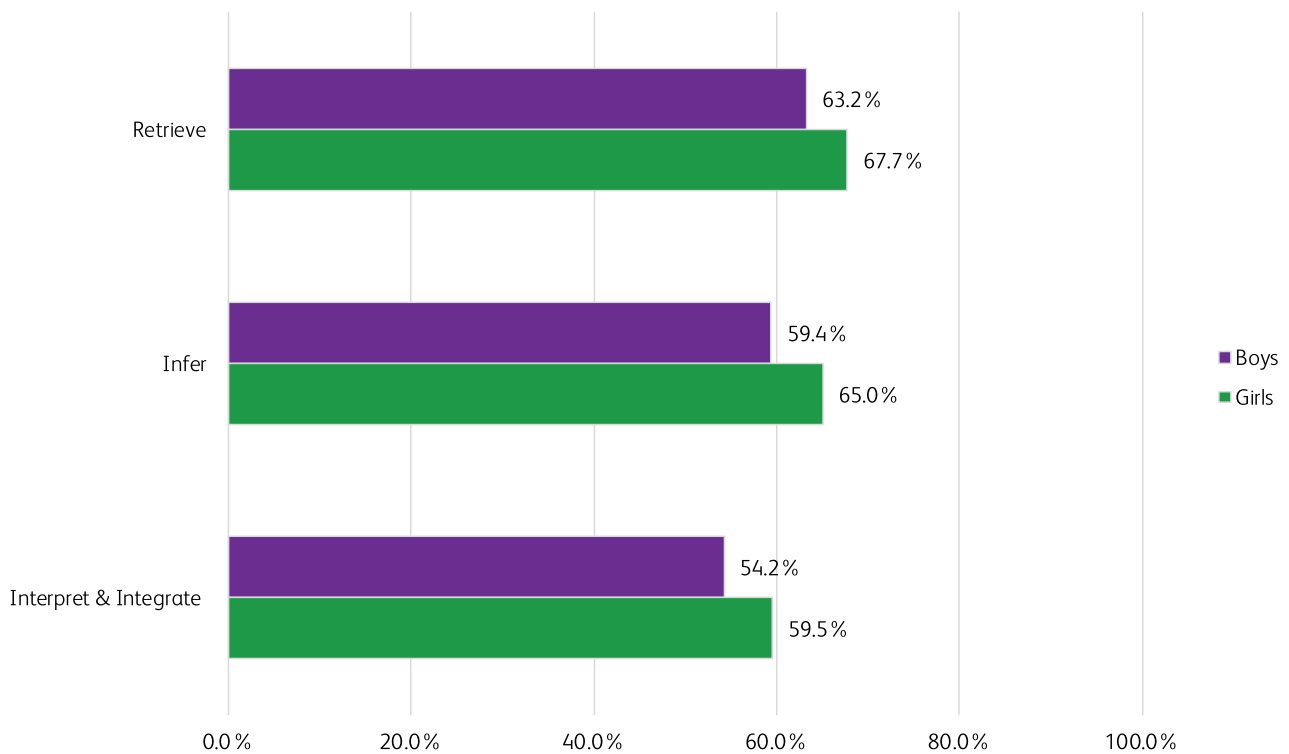


Figure 5.2. Mean percent correct scores by English Reading process skill and gender, NAMER 2021

An examination of the scale scores by gender and process skill reveals similar patterns. As shown in table 5.3, girls had higher mean scores than the other two gender groups in all three reading process skill areas. However, the performance difference between girls and boys is statistically significant only for the Infer subscale. Girls achieved a mean score of 264.9 on this subscale, outperforming boys (257.1) by almost 8 score points. Although girls’ mean scores on the Retrieve and Interpret & Integrate subscales were also higher than boys, neither of these gaps is statistically significant. The effect sizes of the differences in scores between boys and girls in all three process skill areas are small.

Table 5.3. Mean scale scores by English Reading process skill and gender, NAMER 2021

| | Boys | Girls | Boys - Girls |
|----------------------------------|-------------------------|--------------|--------------|
| | <i>Mean scale score</i> | | <i>d</i> |
| Retrieve | 254.6 | 261.2 | 0.14 |
| Infer | 257.1 | 264.9 | 0.16 |
| Interpret & Integrate | 257.2 | 263.1 | 0.13 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Girls’ scores in bold are statistically significantly different from the corresponding boys’ score.

Proficiency levels

As explained in Chapter 4, proficiency levels represent clusters of skills, and describe the types of tasks which pupils at different levels of performance can consistently complete. In 2021, the percentage of girls achieving at or below proficiency Level 1 was 20.3 %, which is statistically significantly lower than the percentage of boys achieving at the lower levels (28.3 %). The percentage of girls achieving at or above Level 3 (47.2 %) was higher than the corresponding percentage of boys (41.2 %), although this difference is not statistically significant. At level 2, the percentages of girls and boys were 32.5 % and 30.5 %, respectively. Of pupils who selected Other as their gender, 28 % achieved at or below level 1, 20.6 % achieved at level 2, and 51.4 % achieved at or above level 3.

Table 5.4. Percentages of pupils achieving at each proficiency level on the overall reading scale by gender, NAMER 2021

| | Boys | Girls |
|---------------|------|-------------|
| | % | % |
| Below Level 1 | 6.6 | 5.3 |
| Level 1 | 21.7 | 15.0 |
| Level 2 | 30.5 | 32.5 |
| Level 3 | 30.8 | 34.1 |
| Level 4 | 10.4 | 13.1 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Girls' percentages in bold are statistically significantly different from the corresponding boys' percentage.

Trends in Second class English reading performance

This section compares pupil performance in 2021 against performance in 2014, in terms of gender, on the overall reading scale, as well as by content area and process skill. Gender differences in achievement across cycles are also considered with reference to proficiency levels.

Trends in performance on overall reading scale, by content area and process skill

As noted in Chapter 4, trend comparisons based solely on percent correct scores may be misleading, because they cannot take into account any changes in the difficulty of the test that may arise from the replacement of items from one cycle to the next. Scale scores generated using the IRT framework take into account such changes in difficulty and may therefore be considered a more reliable and unbiased basis from which to make cross-cycle comparisons. For interest, the percent correct scores for boys and girls across NAMER 2009, 2014 and 2021 can be found in the Appendix (Table A5.5).

In terms of scale scores, boys' overall performance on the English reading assessment dropped from 260.4 in 2014 to 256.8 in 2021. Boys achieved slightly lower scores in 2021 than in 2014 across all content areas and process skills (except for the Interpret & Integrate subscale, on which boys achieved the same score in both cycles). Despite this pattern of slightly lower performance in 2021, none of the decreases in scores are statistically significant and the relevant effect sizes are small.

Table 5.5. Boys' mean scale scores by Reading component, process skill and year

| | 2009 | 2014 | 2021 | 2014-2021 |
|-----------------------|-------------------------|-------|-------|-----------|
| | <i>Mean scale score</i> | | | <i>d</i> |
| Content area | | | | |
| Vocabulary | 245.8 | 263.3 | 260.8 | 0.05 |
| Comprehension | 242.6 | 258.1 | 254.8 | 0.07 |
| Process skill | | | | |
| Retrieve | 242.8 | 256.8 | 254.6 | 0.05 |
| Infer | 244.6 | 258.8 | 257.1 | 0.04 |
| Interpret & Integrate | 243.5 | 257.2 | 257.2 | 0.00 |
| Overall | 243.4 | 260.4 | 256.8 | 0.07 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Girls showed a similar pattern of performance to boys. Scale scores on the overall reading scale and across all content areas and process skills were slightly lower in 2021 than in 2014. As shown in Table 5.6, girls in Second class in 2014 had a mean English reading score of 267.5, and their performance dropped to 265.1 in 2021. The largest mean differences in girls' performance between cycles were on the Comprehension and Retrieve subscales. However, none of the differences in girls' performance between 2014 and 2021 are statistically significant, and the effect sizes of these differences are small.

Table 5.6. Girls' mean scale scores by Reading component, process skill and year

| | 2009 | 2014 | 2021 | 2014-2021 |
|-----------------------|-------------------------|-------|-------|-----------|
| | <i>Mean scale score</i> | | | <i>d</i> |
| Content area | | | | |
| Vocabulary | 254.4 | 266.6 | 264.6 | 0.04 |
| Comprehension | 257.7 | 266.8 | 263.9 | 0.06 |
| Process skill | | | | |
| Retrieve | 257.6 | 264.1 | 261.2 | 0.07 |
| Infer | 255.7 | 267.6 | 264.9 | 0.06 |
| Interpret & Integrate | 256.8 | 265.0 | 263.1 | 0.04 |
| Overall | 256.9 | 267.5 | 265.1 | 0.05 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Trends in performance - proficiency levels

In NAMER 2021, the proportion of Second class boys performing at or below Level 1 (28.3 %) on the overall English reading scale was almost five percentage points higher than the corresponding value in 2014 (23.4 %). The proportion of girls performing at or below Level 1 also increased, albeit very slightly, from 20 % in 2014 to 20.3 % in 2021. The proportion of boys performing at the lowest proficiency levels in 2021 (at or below Level 1) was eight percentage points higher than the proportion of girls performing at the lowest levels.

There was a small decrease in the proportion of boys performing at Level 2, from NAMER 2014 (33.8 %) to NAMER 2021 (30.5 %). The proportion of girls performing at Level 2 increased by 0.6 percentage points, from 31.9 % in 2014 to 32.5 % in 2021.

At the upper proficiency levels, 41.2 % of boys in NAMER 2021 performed at or above Level 3; the corresponding value from NAMER 2014 is 42.9 %. A similarly small gap was observed for higher-achieving girls across cycles - 47.2 % of girls in 2021 performed at or above Level 3, while 48.1 % of their NAMER 2014 counterparts performed at the same levels.

Table 5.7. Percentages of pupils performing at each proficiency level on the overall reading scale by gender and year

| | Boys | | | Girls | | |
|----------------------|------|------|------|-------|------|------|
| | 2009 | 2014 | 2021 | 2009 | 2014 | 2021 |
| Below Level 1 | 12.5 | 6.5 | 6.6 | 6.9 | 3.8 | 5.3 |
| Level 1 | 27.8 | 16.9 | 21.7 | 22.6 | 16.2 | 15.0 |
| Level 2 | 29.5 | 33.8 | 30.5 | 30.5 | 31.9 | 32.5 |
| Level 3 | 22.5 | 31.3 | 30.8 | 27.7 | 32.7 | 34.1 |
| Level 4 | 7.7 | 11.6 | 10.4 | 12.2 | 15.4 | 13.1 |

Note. 2021 percentages in bold are statistically significantly different from the corresponding 2014 percentage.

Sixth class mathematics performance

This section presents Sixth class pupils' mathematics performance in NAMER 2021, and compares boys' and girls' performance. First, pupil achievement is considered in terms of percent correct and scale scores on the overall mathematics scale, as well as by content area and process skill. Then, achievement is outlined with reference to proficiency levels.

Performance on overall mathematics scale, by content area and process skill

Figure 5.3 presents percent correct scores for pupils in all gender categories, on the overall mathematics scale and on the content area subscales. On the overall mathematics scale in 2021, 61.2 % of test items were answered correctly by boys. The corresponding figure for girls was 57.8 %. Pupils who selected 'Other' as their gender answered 56.7 % of items correctly. The Items on the Number & Algebra subscale were the easiest for boys and girls, whose percent correct scores on this subscale were 65.0 % and 61.4 %, respectively. The highest percent correct score for pupils who identified their gender as 'Other' was observed on the Shape & Space subscale (62.9 %). Items on the Measures subscale were the most difficult for all gender groups, with percent correct ranging from 39.6 % (Other) to 47.4 % (Boys).

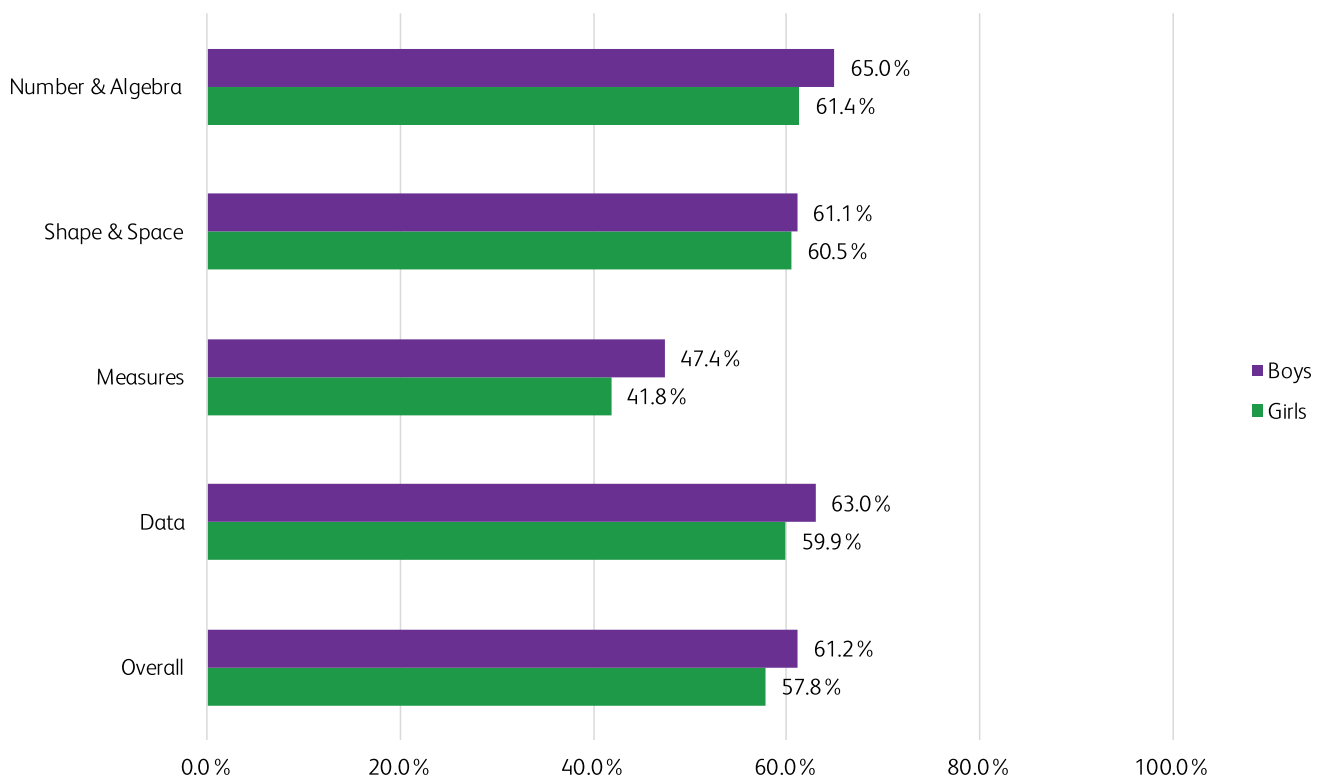


Figure 5.3. Mean percent correct scores by mathematics content area and gender, NAMER 2021

The scale scores in Table 5.8 show that boys outperformed girls on the overall mathematics scale by 7.8 score points, which is a statistically significant difference. Boys also achieved statistically significantly higher scores than girls on the Number & Algebra and Measures subscales.

No statistically significant gender differences were observed in performance on the Shape & Space or Data subscales. The effect sizes for the performance gaps between boys and girls are small, for the overall mathematics scale and for all of the content subscales.

Pupils who selected 'Other' as their gender achieved a mean score of 253.9 on the overall scale, and performed at a similar level on the Measures subscale (253.7), the Number & Algebra subscale (251.4) and the Shape & Space subscale (255.4). Their performance on the Data subscale was slightly higher (258.8).

Table 5.8. Mean scale scores in Mathematics by content area and gender, NAMER 2021

| | Boys | Girls | Boys - Girls |
|-----------------------------|-------------------------|--------------|--------------|
| | <i>Mean scale score</i> | | <i>d</i> |
| Number & Algebra | 263.2 | 255.2 | 0.16 |
| Shape and space | 257.3 | 255.5 | 0.04 |
| Measures | 262.3 | 253.4 | 0.17 |
| Data | 259.0 | 255.7 | 0.07 |
| Overall | 264.6 | 256.8 | 0.16 |

Note. N(Boys) = 2618, N(Girls) = 2697, N(Other) = 65. Girls' scores in bold are statistically significantly different from the corresponding boys' score.

Figure 5.4 shows the proportion of items answered correctly by each gender group on each mathematics process skill subscales in 2021. Items on the Apply & Problem Solve scale were the most difficult for all gender groups with the percent correct ranging across groups from 49.8% (Other) to 55.7% (Boys). Items on the Integrate & Connect subscale were the easiest for all gender groups. Girls answered 65.3% of these items correctly, while boys achieved a percent correct score of 67.3%. Pupils who selected 'Other' as their gender answered 68.3% of Integrate & Connect items correctly. Across all but one of the process subscales, boys answered the highest proportion of items correctly. The exception was the Integrate & Connect subscale, on which pupils in the 'Other' category achieved a slightly higher percent correct than boys.

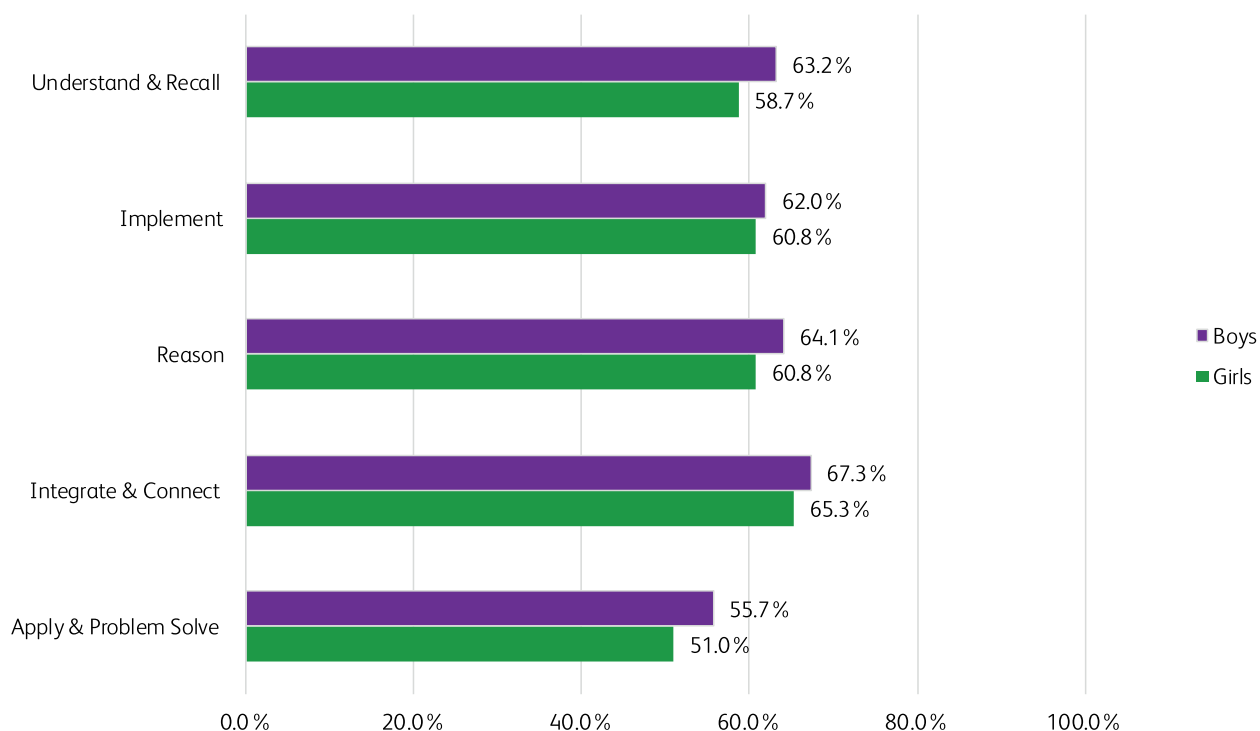


Figure 5.4 Mean percent correct scores by mathematics process skill and gender, NAMER 2021

Table 5.9 displays the scale scores for each gender group across the process skills, as well as the effect sizes for the differences between boys and girls. On the Understand & Recall subscale, boys outperformed girls by 7 scale score points and this difference is statistically significant. Similarly, on the Reason and Apply & Problem Solve subscales, boys achieved scale scores that are statistically significantly higher than girls by 6.9 and 8.4 scale score points, respectively. While boys also achieved higher scale scores than girls on the Implement and Integrate & Connect subscales, these differences are not statistically significant. For all process skill subscales, the effect sizes for the differences between boys and girls are small.

Table 5.9. Mean scale scores by mathematics process skill and gender, NAMER 2021

| | Boys | Girls | Boys - Girls |
|-----------------------|-------------------------|--------------|--------------|
| | <i>Mean scale score</i> | | <i>d</i> |
| Understand & Recall | 261.5 | 254.5 | 0.15 |
| Implement | 261.3 | 258.6 | 0.05 |
| Reason | 263.9 | 257.0 | 0.14 |
| Integrate & Connect | 259.0 | 255.6 | 0.07 |
| Apply & Problem Solve | 262.6 | 254.2 | 0.17 |

Note. N(Boys) = 2618, N(Girls) = 2697, N(Other) = 65, except Integrate & Connect N(Boys) = 1958, N(Girls) = 1994, N(Other) = 50. Girls' scores in bold are statistically significantly different from the corresponding boys' score.

Proficiency levels

Table 5.10 displays the percentages of pupils achieving at each proficiency level for all gender groups. Six percent of boys and 6.3 % of girls achieved below Level 1 on the assessment. For the group that indicated their gender as 'Other', the corresponding figure was 5.5 %. At the opposite end of the scale, 15.6 % of boys reached the highest level (Level 4), while 12.8 % of girls achieved at this level. Almost 10 % of pupils in the 'Other' category achieved at Level 4. There is a statistically significant difference between boys and girls performing at Level 1, where the proportion of girls is significantly higher than the proportion of boys. At each of the other proficiency levels, there are no statistically significant gender differences.

When the proficiency levels are collapsed into categories (At or below Level 1, At or above Level 3), the percentage of girls performing at the lower levels (at or below Level 1) was 31.2 % and this is statistically significantly higher than the proportion of boys (23.1 %). At the higher levels, 45.2 % of boys achieved at or above Level 3, which is statistically significantly higher than the corresponding figure for girls (37.9 %).

Table 5.10. Percentages of pupils achieving at each proficiency level on the overall mathematics scale by gender, NAMER 2021

| | Boys | Girls |
|---------------|------|-------------|
| | % | % |
| Below Level 1 | 6.0 | 6.3 |
| Level 1 | 17.1 | 24.9 |
| Level 2 | 31.7 | 30.9 |
| Level 3 | 29.6 | 25.2 |
| Level 4 | 15.6 | 12.8 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Girls' percentages in bold are statistically significantly different from the corresponding boys' percentage.

Trends in Sixth class mathematics performance

This section compares pupil performance in 2021 against performance in 2014, in terms of gender, on the overall mathematics scale, as well as by content area and process skill. Gender differences in achievement across cycles are also considered with reference to proficiency levels.

Trends in performance on overall mathematics scale, by content area and process skill

As noted earlier in this chapter, trend comparisons based solely on percent correct scores may be misleading, as they do not account for any changes in the difficulty of the test that may result from items being retired and replaced across cycles. Scale scores generated using the IRT framework take into account such changes in difficulty and may therefore be considered a more reliable basis for cross-cycle comparisons. For interest, the percent correct scores for boys and girls across NAMER 2009, 2014 and 2021 can be found in the Appendix (Table A5.16).

Table 5.11 shows the mean scale scores for boys across the last three cycles of NAMER. Many of the subscale scores were slightly lower in 2021 than in 2014. However, boys’ mean score on the overall mathematics scale was slightly higher in 2021 (264.6) than in 2014 (263.7).

The largest difference in boys’ mean performance between 2014 and 2021 is in the content area of Shape & Space. In 2014 boys achieved a score of 262.4 on Shape & Space, while in 2021 boys’ mean score was 257.3, a difference of 5.1 score points. Boys’ performance was slightly lower on the process skill subscales of Implement and Apply & Problem Solve in 2021, compared with 2014. Meanwhile, boys performed slightly higher in 2021 than 2014 on the content area of Number and Algebra and the process skills of Understand & Recall, Reason and Integrate & Connect. Performance on the content areas Measures and Data was very similar in 2021 and 2014. None of the differences in boys’ performance between 2014 and 2021 are statistically significant, and the effect sizes of these differences are small.

Table 5.11 Boys’ mean scale scores by mathematics content area, process skill and year

| | 2009 | 2014 | 2021 | 2021-2014 |
|-----------------------|-------|-------------------------|-------|-----------|
| Content Area | | <i>Mean scale score</i> | | <i>d</i> |
| Number & Algebra | 252.6 | 262.6 | 263.2 | 0.01 |
| Shape and space | 250.8 | 262.4 | 257.3 | 0.11 |
| Measures | 254.2 | 262.1 | 262.3 | 0.00 |
| Data | 251.3 | 259.7 | 259.0 | 0.01 |
| Process Skill | | | | |
| Understand & Recall | 251.5 | 260.4 | 261.5 | 0.02 |
| Implement | 251.1 | 262.3 | 261.3 | 0.02 |
| Reason | 251.6 | 261.7 | 263.9 | 0.04 |
| Integrate & Connect | 252.3 | 257.9 | 259.0 | 0.02 |
| Apply & Problem Solve | 254.3 | 263.4 | 262.6 | 0.01 |
| Overall | 252.8 | 263.7 | 264.6 | 0.02 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

As shown in Table 5.12, girls’ overall performance on the mathematics assessment dropped from 260.0 in 2014 to 256.8 in 2021. Girls achieved slightly lower scale scores on all the content area and process skill subscales in 2021, compared with 2014. This decrease is statistically significant only for the Shape & Space subscale. The corresponding effect sizes are also small.

Table 5.12 Girls' mean scale scores by mathematics content area, process skill and year

| | 2009 | 2014 | 2021 | 2021-2014 |
|-----------------------|-------|-------------------------|--------------|-----------|
| Content Area | | <i>Mean scale score</i> | | <i>d</i> |
| Number & Algebra | 247.2 | 259.2 | 255.2 | 0.08 |
| Shape and space | 249.1 | 263.2 | 255.5 | 0.17 |
| Measures | 245.5 | 255.4 | 253.4 | 0.04 |
| Data | 248.6 | 258.4 | 255.7 | 0.05 |
| Process Skill | | | | |
| Understand & Recall | 248.5 | 260.1 | 254.5 | 0.12 |
| Implement | 248.9 | 263.2 | 258.6 | 0.09 |
| Reason | 248.2 | 261.4 | 257.0 | 0.09 |
| Integrate & Connect | 247.6 | 257.0 | 255.6 | 0.03 |
| Apply & Problem Solve | 245.4 | 255.6 | 254.2 | 0.03 |
| Overall | 247.0 | 260.0 | 256.8 | 0.07 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Trends in performance – proficiency levels

In NAMER 2021, the percentage of boys (15.6 %) and girls (12.8 %) performing at the highest level (Level 4) was lower than in 2014, when 17 % of boys and 13.2 % of girls achieved at this level. At Level 3, the second highest level of performance, there was a small increase in the proportion of boys performing at this level between 2014 (26.6 %) and 2021 (29.6 %). There was a corresponding decrease in the proportion of girls achieving Level 3, with 32 % achieving Level 3 in 2014 and 30.9 % in 2021.

The proportion of boys performing below Level 1 was 6.7 % in 2014 and dropped slightly to 6.0 % in 2021. The proportion of girls performing below Level 1 increased from 4.6 % in 2014 to 6.3 % in 2021. There was a small decrease in the proportion of boys performing at Level 1 in 2021 (17.1 %), when compared with 2014 (19.1 %). The trend was different for girls performing at Level 1, with the proportion increasing slightly from 22.5 % in 2014 to 24.9 % in 2021. At Level 2, the proportion of boys achieving this level increased by one percentage point from 30.7 % in 2014 to 31.7 % in 2021. The proportion of girls at Level 2 dropped slightly from 32 % in 2014 to 30.9 % in 2021. However, none of these differences between 2014 and 2021 are statistically significant.

Table 5.13 Percentages of pupils performing at each proficiency level on the overall mathematics scale by gender and year

| | Boys | | | Girls | | |
|----------------------|------|------|------|-------|------|------|
| | 2009 | 2014 | 2021 | 2009 | 2014 | 2021 |
| Below Level 1 | 9.2 | 6.7 | 6.0 | 9.8 | 4.6 | 6.3 |
| Level 1 | 23.6 | 19.1 | 17.1 | 25.6 | 22.5 | 24.9 |
| Level 2 | 29.4 | 30.7 | 31.7 | 29.8 | 32.0 | 30.9 |
| Level 3 | 26.3 | 26.6 | 29.6 | 24.5 | 27.7 | 25.2 |
| Level 4 | 11.4 | 17.0 | 15.6 | 10.3 | 13.2 | 12.8 |

Note. 2021 percentages in bold are statistically significantly different from the corresponding 2014 percentage.

Summary

In 2021, for the first time, pupils had the option to select one of three gender categories – Boy, Girl and Other. Results are only compared for boys and girls due to the very small number of pupils who selected Other.

English reading

Similarly to 2014, girls in Second class had a statistically significantly higher overall English reading mean score than boys in NAMER 2021. Girls also outperformed boys on the Comprehension subscale, but boys and girls performed similarly to each other on the Vocabulary subscale. In terms of the process skill subscales, the only significant difference in performance between boys and girls in 2021 is on the Infer subscale; girls and boys performed similarly to each other on the Retrieve and Interpret & Integrate subscales.

The percentage of girls achieving at the lower proficiency levels was statistically significantly lower than the corresponding percentage of boys. A greater percentage of girls than boys achieved at the higher proficiency levels but this difference is not statistically significant. There are no statistically significant changes in the proportion of boys and girls performing at each proficiency level in 2021, compared with 2014.

Mathematics

In NAMER 2021, boys performed at a higher level than girls overall and on all subscales. The differences are statistically significant on the overall mathematics scale and on the Measures and Number & Algebra subscales. Boys also statistically significantly outperformed girls on the process subscales of Understand & Recall, Reason and Apply & Problem Solve. Girls and boys performed similarly to each other on the Shape & Space, Data, Implement and Integrate & Connect subscales.

In terms of proficiency levels, the proportion of boys achieving at or above Level 3 is statistically significantly higher than the proportion of girls achieving at these highest levels. Conversely, the proportion of girls achieving at the lowest proficiency levels is statistically significantly higher than the corresponding percentage of boys.

There were slight decreases in the proportions of boys and girls performing at the highest proficiency levels in 2021 when compared with performance in 2014, but these decreases are not statistically significant. There was also a slight increase in the number of girls performing at the lowest proficiency levels, but this is not statistically significantly different from the proportion of girls performing at these levels in 2014.

CHAPTER 6

Summary and Conclusions

In this chapter, the performance of pupils in the 2021 National Assessments is summarised, and key conclusions are drawn. The context report, which will be published in 2023, will seek to interpret the findings in greater detail, drawing on the outcomes of questionnaires completed by participating pupils, their class teachers and their school principals.

Sample and tests

The NAMER 2021 tests were administered by class teachers, under the supervision of a nominated school coordinator, to a representative sample of over 10,000 pupils in Second and Sixth classes in 188 primary schools in May 2021. Inspectors from the Department of Education supervised the administration of testing in 20% of schools.

The pupils that participated in the tests had experienced two months of remote teaching and learning in January and February 2021, returning to face-to-face teaching in school in March 2021.

Second class pupils took the English reading assessment only, and Sixth class pupils took the mathematics test only. There were four forms of the Second class reading test and eight forms of the Sixth class mathematics test. These were secure curriculum-based instruments which were originally developed for the 2009 National Assessments (Eivers et al., 2010). They were updated for the 2014 National Assessments and had undergone further updates for NAMER 2021. A small number of new test items were included in the test booklets in each cycle, to replace items which had been released, items that were not functioning as expected or items that had simply become outdated. At Second class, 91.7% of pupils took the test, 1.8% were exempt, and 6.4% were absent from school on the day of testing. At Sixth class, 91.4% of pupils took the test, 1.2% were exempt, and 7.4% were absent on the day of testing. In addition to these tests, pupils, their class teachers and school principals completed questionnaires.

English reading

Second class pupils' overall performance on English reading in NAMER 2021 was similar to their counterparts' in 2014. In 2021, the mean score on the overall reading scale was 260.8, which was lower than the overall mean score in 2014 (264.0), although the difference in scores is not statistically significant.

Similarly, no notable change was observed in the 2021 mean scores for the Vocabulary and Comprehension subscales, or for any of the process skill subscales (Retrieve, Infer, Interpret & Integrate) when compared with the corresponding scores from NAMER 2014. The stability of these results across cycles is reflected in small effect sizes, ranging from 0.02 to 0.07, which indicate that any changes in scores between 2014 and 2021 do not represent meaningful differences in pupil performance.

There are no statistically significant differences between 2014 and 2021 in scores at key percentile points on the overall reading scale, or in scores on the Vocabulary or Comprehension subscales, at key percentile points between 2014 and 2021. However, there are differences in scores at key percentile points across cycles on some of the process skill subscales; pupils' scores at the 75th percentile on the Retrieve and Interpret & Integrate subscales are statistically significantly lower than the corresponding in 2014 scores, while the scores at the 90th percentile on the Interpret & Integrate subscale in 2021 is statistically significantly higher than the corresponding score in 2014. These results show no discernible pattern of change in pupil performance, since scores in 2021 are neither consistently higher nor consistently lower than the corresponding scores in 2014.

The percentages of Second class pupils performing at the different proficiency levels in reading in NAMER 2021 were broadly similar to the corresponding percentages in 2014. The percentage of pupils performing at or above Level 3 in 2021 (44.1 %) was very slightly lower than in 2014 (45.5 %), and the percentage of pupils performing at or below Level 1 in NAMER 2021 (24.4 %) was slightly higher than in 2014 (21.7 %). However, these differences are not statistically significant, which indicates that the proportions of high- and low-performing pupils are similar in both cycles.

The revised targets set in the Interim Review of the National Strategy for Literacy and Numeracy (Department of Education and Skills, 2017) sought an increase to 50 % of pupils performing at or above Level 3, and a reduction to 20 % of pupils performing at or below Level 1. However, the percentage of pupils performing at or below Level 1 in 2021 was statistically significantly higher than the target, and the percentage of pupils performing at or above Level 3 in 2021 was statistically significantly lower than the target. Therefore, the revised targets were not met at either end of the achievement distribution. Despite this, it is worth noting that the percentages of pupils achieving at or below Level 1 and at or above Level 3 in 2021 meet the original performance targets set in the 2011 Strategy for Literacy and Numeracy.

Mathematics

Overall performance on the mathematics assessment at Sixth class in 2021 was broadly similar to performance in 2014. This mean overall scale score in 2021 was 260.5, about one scale score point less than the overall mean in 2014 (261.7). The difference in performance between 2014 and 2021 is not statistically significant and the corresponding effect size ($d = 0.02$) is small, suggesting that the overall mean scale score in 2021 is not meaningfully different from the overall mean scale score in 2014.

Performance on the content and process subscales also did not change significantly between 2014 and 2021. Only performance on the content subscale Shape & Space (256.4) was statistically significantly lower than the score in 2014 (268.2) with an effect size of 0.14.

Overall, there are no statistically significant differences in pupils' achievement in Mathematics at key percentile points between 2014 and 2021. However, there are differences in scores at key percentile points across cycles on two subscales; pupils' scores at the 50th and 75th percentiles on the Shape & Space subscale and their score at the 75th percentile on the Integrate & Connect subscale, are statistically significantly lower than the corresponding 2014 scores, while pupils' scores at the 25th and 90th percentiles on the Integrate & Connect subscale in 2021 are statistically significantly higher than the corresponding scores in 2014. These results show no discernible pattern of change in pupil mathematics performance at key percentile points, since scores in 2021 are neither consistently higher nor consistently lower than the corresponding scores in 2014.

The percentages of Sixth class pupils performing at each proficiency level in Mathematics have remained relatively stable between 2014 and 2021. The percentage of pupils performing at or above Level 3 in 2021 (41.4 %) was very slightly lower than in 2014 (42.1 %), and the percentage of pupils performing at or below Level 1 in NAMER 2021 (27.3 %) was very slightly higher than in 2014 (26.5 %). These differences are not statistically significant, which indicates that the proportions of high- and low-performing pupils in Mathematics are similar in both cycles.

The revised targets set in the Interim Review of the National Strategy for Literacy and Numeracy (Department of Education and Skills, 2017) sought an increase to 50 % of pupils performing at or above Level 3, and a reduction to 20 % of pupils performing at or below Level 1. However, the percentage of pupils performing at or below Level 1 in 2021 was statistically significantly higher than the target, and the percentage of pupils performing at or above Level 3 in 2021 was statistically significantly lower than the target, indicating that the revised targets were not met at either end of the achievement distribution.

Gender differences

In NAMER 2021, for the first time, pupils were given three options on the questionnaire to identify their gender – Girl, Boy and Other. Comparisons were only drawn between girls’ and boys’ results, as only a very small number of pupils selected Other and therefore, statistical comparisons with this group would not be reliable.

Gender and English reading

Second class pupils’ English reading performance followed a similar pattern to their counterparts in 2014, in terms of gender differences, with girls tending to perform more highly than boys. Like in 2014, girls in 2021 outperformed boys on the overall reading scale; girls’ score (265.1) was about eight score-points higher than boys’ (256.8). This performance gap is statistically significant and has an effect size of 0.17. Girls also performed statistically significantly higher than boys on the Comprehension subscale, but there is no statistically significant performance gap between girls and boys on the Vocabulary subscale.

In terms of process skills, girls’ mean scores in 2021 were higher than boys’ on all three subscales, but the performance difference is statistically significant only for the Infer subscale, on which girls achieved a mean score of 264.9, almost 8 score-points higher than boys’ mean score (257.1). Effect sizes of the differences in scores between boys and girls are small on all three process skill subscales.

In NAMER 2021, the proportion of girls performing at or below proficiency Level 1 (20.3 %) is statistically significantly lower than the proportion of boys performing at this level (28.3 %). At the upper end of the achievement distribution, a greater percentage of girls (47.2 %) than boys (41.2 %) performed at or above Level 3, but this difference is not statistically significant.

Analysis of trends in English reading performance by gender showed that Second class boys achieved lower scores in 2021 than in 2014, overall and across all content areas and process skills (except for the Interpret & Integrate subscale, on which boys achieved the same score in 2014 and 2021). However, none of the decreases in scores is statistically significant and the relevant effect sizes are small. The proportion of boys achieving at or below proficiency Level 1 in NAMER 2021 (28.3 %) was almost five percentage points higher than the corresponding value in 2014 (23.4 %), while the proportion of boys achieving at or above Level 3 was similar in 2021 (41.2 %) and 2014 (42.9 %). These changes in proportions of boys performing at the highest and lowest proficiency levels are not statistically significant.

The pattern of performance across cycles is similar for girls, who had lower scores in 2021 than in 2014, on the overall reading scale and across all content areas and process skills. Similarly to boys, none of the differences in girls’ performance between 2014 and 2021 are statistically significant and the associated effect sizes are small. The proportion of girls performing at the lowest proficiency levels in 2021 (at or below Level 1) showed only a marginal increase from 20 % in 2014 to 20.3 % in 2021. Similarly, small changes were observed in the proportion of girls achieving at the highest proficiency levels (at or above Level 3) between 2014 (48.1 %) and 2021 (47.2 %). None of the changes observed in the proportion of girls achieving at various proficiency levels are statistically significant.

Gender and Mathematics

Sixth class pupils’ performance on the mathematics assessment followed a similar pattern to 2014, with boys on average performing at a higher level than girls. As in 2014, boys in 2021 had a higher mean score than girls on the overall mathematics scale; boys’ score (264.6) was 7.8 score points higher than girls’ (256.8). This performance gap is statistically significant and has an effect size of 0.16.

Boys’ average performance was also statistically significantly higher than girls on the Number & Algebra and Measures subscales, but the performance gaps on the Data and Shape & Space subscales are not

statistically significant. The effect sizes for the performance gaps on the content subscales are all small.

Overall, boys' mean scores on the process subscales were higher than girls mean scores. On average boys performed statistically significantly higher than girls on the Understand & Recall subscale by 7 scale score points. Similarly, boys' average achievement on the Reason and Apply & Problem Solve subscales is statistically significantly higher than girls by 6.9 and 8.4 scale score points, respectively. For all process skill subscales, the effect sizes for the differences between boys and girls are small.

The proportion of boys performing at or below proficiency Level 1 (23.1 %) is statistically significantly smaller than the proportion of girls at this level (31.2 %). The proportion of boys performing at or above Level 3 (45.2 %) is statistically significantly larger than the corresponding proportion of girls (37.9 %).

Trends in mathematics performance by gender showed that girls in Sixth class achieved slightly lower scores in 2021 when compared to 2014 to the corresponding proportion 260.0 in 2014 to 256.8 in 2021. The same trend in performance for girls was seen in all content and process subscales. However, this decrease in performance is only statistically significant for the Shape & Space subscale. The corresponding effect sizes are again small.

There was a slight increase in the proportion of girls at the lowest levels of performance (31.2 %), but this is not statistically significantly different to performance in 2014 (27.1 %). The proportion of girls at or above Level 3 was slightly lower in 2021 (37.9 %) when compared to 2014 (40.9 %), but again this is not statistically significant.

The pattern of performance for boys, with many of the subscale scores, is slightly lower in 2021 than in 2014. However, boys' mean score on the overall mathematics scale was slightly higher in 2021 (264.6) than in 2014 (263.7). There were also slightly fewer boys performing at or below Level 1 in 2021 (23.1 %) when compared to 2014 (25.8 %)

Overall, there was a slight increase in the number of boys achieving the highest levels of achievement in 2021 (45.2 %) when compared with performance in 2014 (43.6 %), but the difference was small and is not statistically significantly different to 2014. None of the changes observed in the proportion of boys achieving at various proficiency levels are statistically significant.

Conclusions

This section situates the results of NAMER 2021 in the context in which the study took place. That is, performance in NAMER 2021 is discussed in relation to trends in international studies, the National Literacy and Numeracy Strategy, the introduction of the new PLC, and the COVID-19 pandemic.

Performance on English reading and Mathematics

Drawing comparisons between results from National Assessments and those from international studies (e.g., PIRLS, TIMSS, PISA) can offer further insights into trends in pupil performance. For example, contextualising the unprecedented gains in English reading and mathematics achievement seen in NAMER 2014 by examining the international studies from that time can support the validation of these gains. The replication of similar patterns of achievement in international studies strengthens the likelihood that the gains observed were, in fact, increases in educational attainment.

Since NAMER 2014, the following international assessments have been conducted at primary level in Ireland: PIRLS 2016 and 2021, TIMSS 2015 and 2019 and PISA 2015 and 2018. At the time of writing, PISA 2022 is underway, and Mathematics is the major domain.

Consistent with the statistically significant gains in reading and mathematics achievement observed in NAMER 2014, there were corresponding large increases in performance in English reading in PIRLS between 2011 (552) and 2016 (567) (Eivers et al., 2017; Eivers & Clerkin, 2012). The observation of similar large gains in reading achievement in PIRLS 2016 suggests that the gains seen in NAMER 2014 reflect a true improvement in English reading standards in Irish primary schools over that period.

Both PISA 2018 and NAMER 2021 showed a slight decline in mean scores in English reading since the last cycles (C. McKeown et al., 2019). However, this slight decline is not statistically significant in either NAMER 2021 or PISA 2018, suggesting overall stability in reading performance since 2014. Prior to 2014, performance on the NAMER English reading assessment had only increased significantly on one occasion – among 10-year-olds between 1972 and 1980 (Mulrooney, 1986).

The PIRLS 2021 results show an 11-point increase in performance since 2016. However, due to the disruption of the pandemic, in PIRLS 2021 pupils were tested at the beginning of Fifth class, rather than at the end of Fourth class as in previous rounds of the assessment. This meant that pupils sitting the test were, on average, six months older than their counterparts who were assessed in 2016. Data for Ireland from PIRLS 2016 showed that pupils who were of medium age (within one standard deviation of the average age) achieved significantly higher reading scores, on average, than younger pupils. The extent to which the increase in performance seen in PIRLS 2021 may be due to increased age or the greater length of time having elapsed since periods of remote learning is unclear, but the results from PIRLS 2021 do suggest relative stability in reading performance between 2016 and 2021 (Delaney et al., 2023).

A large increase in mathematics performance was also observed in NAMER 2014 compared to 2009. A similar trend for increased performance in Mathematics was seen TIMSS 2015 (547) compared to 2011 (527) (Clerkin et al., 2016; Eivers & Clerkin, 2012). As in NAMER 2021, subsequent performance in TIMSS 2019 and PISA 2018 has remained relatively consistent with no significant difference in performance compared to previous cycles in 2015 (TIMSS and PISA), suggesting that the gains in performance observed in NAMER 2014 have been maintained (Clerkin et al., 2016; C. McKeown et al., 2019; Perkins & Clerkin, 2020).

Gender gap

In NAMER 2021, statistically significant gender gaps in performance were identified in both Second class English reading and Sixth class Mathematics.

In NAMER 2021, like NAMER 2014, girls in Second class achieved a statistically significantly higher mean score than boys on the reading test overall. This is consistent with PIRLS 2016, PIRLS 2021, PISA 2015 and PISA 2018, where girls significantly outperformed boys on the English reading test (Eivers et al., 2017; Delaney et al., 2023; Shiel et al., 2016; McKeown et al., 2019). It is interesting to note that in PISA 2018 the gender difference in English reading in Ireland was among the smallest gaps across comparison countries.

There was no statistically significant difference in overall mathematics performance between boys and girls in 2014. However, some gender differences observed in the NAMER 2021 Sixth class mathematics assessment are statistically significant, with boys performing significantly higher than girls on the test overall and on the Number & Algebra and Measures content subscales. Boys also significantly outperformed girls on the Understand & Recall, Reason and Apply & Problem Solve process subscales. This represents a shift since NAMER 2014, when boys scored significantly higher on only two subscales: Measures and Apply & Problem Solve.

Although this gender gap is apparent in international mathematics assessments at post-primary level (e.g., PISA 2015 and 2018), the pattern is different in international studies that assess primary level (e.g., TIMSS 2011, 2015 and 2019). In TIMSS 2015 and 2019 the gap in performance between boys and girls is not statistically significant (Clerkin et al., 2016; Perkins & Clerkin, 2020). In PISA 2018, although it is statistically significant, the gender difference in mathematics performance in favour of boys in Ireland was only slightly larger than the OECD average difference (C. McKeown et al., 2019).

COVID-19 pandemic

As with most things over the last few years, the COVID-19 pandemic had a significant impact on the administration of NAMER 2021, most notably on the decision to reduce the scope of the study in 2021 by administering only the English reading assessment at Second class and the mathematics assessment at Sixth class. The decision was also made not to administer the parent/guardian questionnaire to reduce the burden on schools and teachers. Although the decision to reduce the study and break with previous cycles of NAMER was a difficult one due to the loss of important data, it had a positive outcome for the study. The overall participation rate in NAMER 2021 was 91.5 % despite the difficulties faced by schools in administering the study. Reducing the number of assessments administered to each pupil cut down on the number of days needed to administer the assessment, and the gave schools more flexibility in the assignment of teachers and staff to oversee the study and facilitate social distancing.

As mentioned in the introduction, in previous cycles of NAMER, Inspectors from the Department of Education oversaw testing in schools and delivered assessment materials to the schools on the test date, to protect the security of the tests. However, in 2021, in the context of the pandemic, the administration procedure was changed to minimise visits from external personnel to schools. For this cycle, materials were sent directly to schools by post, and Department of Education Inspectors visited a sample of schools (20 %). Reports from schools visited by inspectors suggest that there was strong adherence to the standardised assessment procedures detailed in the administration manuals, and that the assessments were kept secure.

Given the extended period since NAMER 2014 and the cross-sectional design of the study, it is not possible to draw meaningful conclusions on the specific impact of the COVID-19 pandemic on pupil learning from NAMER 2021.

Literacy and Numeracy Strategy

The National Strategy for Literacy and Numeracy, published in 2011, has been a pillar of primary education policy in Ireland for the last ten years, and has set out a roadmap to improve standards in literacy and numeracy through changes in the way teachers are trained (length of B.Ed. and increased focus on literacy and numeracy in continuing professional development [CPD]) and classroom practice (allocation of teaching time, use of formative and standardised assessment, whole school planning, new Primary Language Curriculum). The initial reading and literacy targets set out in the 2011 strategy were met and surpassed in NAMER 2014. The strategy was reviewed in 2015 and new targets were set in 2017 for NAMER 2020/21.

Prior to NAMER 2014, the availability of CPD in literacy and numeracy was increased for teachers. There was also an increase in the allocation of class time to teaching English reading and Mathematics, reporting of standardised test results of individual pupils to their parents, the reporting of aggregated test results to school Boards of Management and to the Department of Education and Skills, and an enhanced emphasis on literacy and numeracy in the context of school self-evaluation and planning. It is thought that at least some of the improvements in performance seen in NAMER 2014 could be attributed to the interventions outlined in the Literacy and Numeracy Strategy. However, not all the interventions outlined in the strategy could be commented on in 2014 as they had not yet been fully implemented.

NAMER 2020/21 offered an opportunity to explore whether the strategies implemented since 2014, such as the expansion of the B.Ed. from three to four years, and the introduction of the PLC to the Junior Classes, had had an effect on teaching and learning in English reading and Mathematics.

The revised PLC replaces the previous Irish and English Language Curricula published in 1999 and was implemented in stages between 2015 and 2019. The implementation of the PLC for Junior classes began in 2015 and took effect, following professional development and training for teachers, in September 2016. Following feedback from implementation of the Curricula at the junior classes, a more streamlined and improved PLC was issued for all classes in 2019. The PLC aims to enhance links between oral language, reading and writing across a range of subject areas and topics in an effort to strengthen pupils' reading

comprehension. The PLC has a stronger focus on learning outcomes and the use of formative assessment than the 1999 Curriculum.

At the time of writing (December 2022), the redeveloped Primary School Mathematics Curriculum is yet to be implemented. It was initially envisaged that the new curriculum would be released in two parts, following on from the approach used for the PLC. A draft specification for Junior classes was published in 2017, undergoing consultation until March 2018. Feedback from the education system on the phased implementation of the PLC led to the decision to extend the original timeline for development of the mathematics curriculum, and to publish the specification for all classes simultaneously. A draft specification for all class levels was released for consultation in March 2022 (National Council for Curriculum and Assessment, 2022), with the consultation running until the end of June. As with the PLC, the draft specification is based on a learning outcomes model and promotes the use of formative assessment. There is also an increased focus on 'maths talk' and on encouraging 'productive disposition', or positive attitudes towards Mathematics.

Similarly to the 1999 Curriculum, the new specification uses the structure of Strands, but instead of process skills, there are four Elements: Understanding & Connecting, Communicating, Reasoning and Applying & Problem Solving. The new curriculum will be accompanied by the Primary Mathematics Toolkit. This will include the key mathematical concepts underlying the learning outcomes, as well as Progression Continua, which will provide descriptions of mathematical development. The NAMER 2014 report observed that the scope for further substantial progress was likely greater in Mathematics than in reading, in particular, in the area of problem solving (Shiel et al., 2014). The next cycle of NAMER will provide a welcome opportunity to observe any potential impacts of the revised Primary Mathematics Curriculum.

As NAMER 2020 had to be rescheduled to 2021 due to the COVID-19 pandemic, seven years had elapsed since the previous round of NAMER in 2014. Given the extended timeframe since the last implementation of the National Assessment, and in particular due to the extended school shutdowns and remote teaching and learning due to the pandemic in Spring 2020 and 2021, it is difficult to draw strong conclusions about the impact of specific aspects of the Literacy and Numeracy strategy implemented since 2014, such as the PLC and the extension of the primary B.Ed, on literacy and numeracy outcomes. However, it is reassuring that results in both English reading at Second class and Mathematics at Sixth class are broadly consistent with performance recorded in the 2014 National Assessments despite the very significant challenges for the education system during the COVID-19 pandemic. In addition, the results for Second class reading are consistent with the PIRLS 2021 outcomes. While the new targets set in 2017 were not met, results in NAMER 2021 were not statistically significantly different from the results seen in 2014.

The National Strategy referred to a need to raise awareness of the importance of digital literacy and to include a measure of digital literacy as part of the National Assessment of English reading. The inclusion of a test of digital reading in PIRLS 2016 (the e-PIRLS assessment) provided initial insights into the performance of pupils in Fourth class on digital reading tasks, and also pointed to infrastructural issues. PISA and TIMSS have also moved to a digital platform, and it is advised that NAMER contain a measure of digital literacy in future rounds.

It would also seem important, both in the context of developing and implementing the revised mathematics curriculum at primary level, and in planning for future National Assessments of Mathematics, to ensure that adequate attention is paid to the use of Information and Communications Technology (ICT) in Mathematics, given that Irish 15-year-olds in PISA do less well on tests of computer-based Mathematics compared with paper-based Mathematics (Perkins et al., 2013).

References

- Assessment Systems Corporation. (2013). *ITEMAN (Version 4.3) [Computer software]*. Author.
- Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology, 24*(3–4), 323–335. <https://doi.org/10.1080/02702710390227297>
- Cain, K., Lemmon, K., & Oakhill, J. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671–681. <https://doi.org/10.1037/0022-0663.96.4.671>
- Clerkin, A., Perkins, R., & Cunningham, R. (2016). *TIMSS 2015 in Ireland: Mathematics and Science in Primary and Post-Primary Schools*. Educational Research Centre. <https://www.erc.ie/wp-content/uploads/2016/11/TIMSS-initial-report-FINAL.pdf>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57–75. <https://doi.org/10.1016/j.jecp.2015.03.005>
- Delaney, E., McAteer, S., Delaney, M., McHugh, G. & O'Neill, B. (2023). *PIRLS 2021: Reading results for Ireland*. Educational Research Centre.
- Department of Education and Science, & National Council for Curriculum and Assessment. (1999a). *Primary School Curriculum: Introduction*. Stationery Office, Government of Ireland. https://www.curriculumonline.ie/getmedia/c4a88a62-7818-4bb2-bb18-4c4ad37bc255/PSEC_Introduction-to-Primary-Curriculum_Eng.pdf
- Department of Education and Science, & National Council for Curriculum and Assessment. (1999b). *Primary School Curriculum: Mathematics*. Stationery Office, Government of Ireland. https://www.curriculumonline.ie/getmedia/9df5f3c5-257b-471e-8d0f-f2cf059af941/PSEC02_Mathematics_Curriculum.pdf
- Department of Education and Skills. (2011). *Literacy and Numeracy for Learning and Life: The National Strategy to Improve Literacy and Numeracy among Children and Young People 2011-2020*. Author. <https://assets.gov.ie/24521/9e0e6e3887454197a1da1f9736c01557.pdf>
- Department of Education and Skills. (2017a). *DEIS Plan 2017: Delivering Equality Of Opportunity In Schools*. Author. <https://www.gov.ie/pdf/?file=https://assets.gov.ie/24451/ba1553e873864a559266d344b4c78660.pdf#page=null>
- Department of Education and Skills. (2017b). *National Strategy: Literacy and Numeracy for Learning and Life 2011-20*. Interim review 2011-16; New targets: 2017-20. Authors. <https://assets.gov.ie/24960/93c455d4440246cf8a701b9e0b0a2d65.pdf>
- Department of Education and Skills. (2019). *Primary Language Curriculum. Government of Ireland*. https://curriculumonline.ie/getmedia/2a6e5f79-6f29-4d68-b850-379510805656/PLC-Documents_English.pdf
- Educational Research Centre. (2008). *Framework for the Assessment of English Reading (NAER 2009)*. Author. https://www.erc.ie/documents/reading_framework2009.pdf
- Educational Research Centre. (2009). *Framework and Test Specifications for the 2009 National Assessment of Mathematics: Second and Sixth Classes*. Author. http://www.erc.ie/documents/maths_framework2009.pdf
- Educational Research Centre. (2019). *New Drumcondra Primary Reading Test: Administration and Interpretation Manual. Levels 3-6 (Paper version)*. Author.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics*. <https://doi.org/10.1137/1.9781611970319>

- Ehri, L. C. (2014). *Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning*. *Scientific Studies of Reading*, 18(1), 5–21. <https://doi.org/10.1080/10888438.2013.819356>
- Eivers, E., & Clerkin, A. (2012). *PIRLS & TIMSS 2011: Reading, Mathematics and Science Outcomes for Ireland*. Educational Research Centre. <https://assets.gov.ie/24993/5fc773438c374aaf9108805ec229d972.pdf>
- Eivers, E., Clerkin, A., Millar, D., & Close, S. (2010). *The 2009 National Assessments: Technical Report*. Educational Research Centre. https://www.erc.ie/documents/na2009_technical_report.pdf
- Eivers, E., Close, S., Shiel, G., Millar, D., Clerkin, A., Gilleece, L., & Kiniry, J. (2009). *The 2009 National Assessments of Mathematics and English Reading*. Department of Education and Skills.
- Eivers, E., Gilleece, L., & Delaney, E. (2017). *Reading Achievement in PIRLS 2016: Initial Report for Ireland*. Educational Research Centre. https://www.erc.ie/wp-content/uploads/2017/12/PIRLS-2016_inital-report-IRL.pdf
- Eivers, E., Shiel, G., Perkins, R., & Cosgrove, J. (2005). *The 2004 National Assessment of English Reading*. Educational Research Centre. https://www.erc.ie/documents/naer04_report_full.pdf
- Foy, P., & LaRoche, S. (2020). *Estimating standard errors in the TIMSS 2019 results*. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). https://timssandpirls.bc.edu/timss2019/methods/pdf/T19_MP_Ch14-estimating-standard-errors.pdf
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Publications, Inc.
- Hart, B., & Risley, T. (1995). *Meaningful Differences in Everyday Experience of Young American Children*. P.H. Brookes.
- IEA. (2022). Help Manual for the IEA IDB Analyzer (Version 5.0). Author.
- Kame'enui, E. J., & Baumann, J. F. (2012). *Vocabulary Instruction: Research to Practice (2nd ed.)*. Guilford Publications.
- Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, 50(2), 151–169. <https://doi.org/10.1002/rrq.99>
- McKeown, C., Denner, S., McAteer, S., & Shiel, G. (2019). *Learning for the Future: The performance of 15-Year-olds in Ireland on Reading Literacy, Science and Mathematics in PISA 2018*. Educational Research Centre. <https://www.erc.ie/wp-content/uploads/2020/07/B23321-PISA-2018-National-Report-for-Ireland-Full-Report-Web-4.pdf>
- McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). *The effects of long-term vocabulary instruction on reading comprehension: A replication*. *Journal of Literacy Research*, 15(1), 3–18. <https://doi.org/10.1080/10862968309547474>
- Mullis, I. V. S., Kennedy, A. M., Martin, M. O., & Sainsbury, M. (2006). *PIRLS 2006 Assessment Framework and Specifications (2nd ed.)*. International Association for the Evaluation of Educational Achievement (IEA). <https://timss.bc.edu/PDF/P06Framework.pdf>
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 Assessment Framework (2nd ed.)*. TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/pirls2016/framework.html>
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). <https://timss2019.org/wp-content/uploads/frameworks/T19-Assessment-Frameworks.pdf>
- Mulrooney, V. J. (1986). National surveys of reading attainment in primary schools in Ireland. In V. Greaney (Ed.), *Irish papers presented at the Fourth European Reading Congress and Tenth RAI Annual Conference* (pp. 187–200). Reading Association of Ireland.

- National Council for Curriculum and Assessment. (2019). *Primary Language Curriculum, Progression Continua*. Government of Ireland. <https://curriculumonline.ie/getmedia/87033bef-fa59-41e0-87e4-4f9126e61821/Progression-Continua.pdf>
- National Council for Curriculum and Assessment. (2022). *Primary Mathematics Curriculum: Draft Specification for Consultation*. Author. https://ncca.ie/media/5370/draft_primary_mathematics_curriculum_specification.pdf
- Nelis, S.M. & Gilleece, L. (2023). *Ireland's 2021 National Assessments of Mathematics and English Reading: A focus on achievement in urban DEIS schools*. Dublin: Educational Research Centre.
- Oakhill, J. V., & Cain, K. (2012). *The precursors of reading ability in young readers: Evidence from a four-year longitudinal study*. *Scientific Studies of Reading*, 16(2), 91–121. <https://doi.org/10.1080/10888438.2010.529219>
- OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. OECD Publishing. <https://doi.org/http://dx.doi.org/10.1787/9789264266490-en>
- OECD. (2019a). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- OECD. (2019b). *PISA 2018 Results: Vol. 1. What Students Know and Can Do (Vol. I)*. PISA, OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- Perkins, R., & Clerkin, A. (2020). *TIMSS 2019: Ireland's Results in Mathematics and Science*. Educational Research Centre. https://www.erc.ie/wp-content/uploads/2020/12/03-ERC-TIMSS-2019-Report_A4_Online.pdf
- Perkins, R., Shiel, G., Merriman, B., Cosgrove, J., & Moran, G. (2013). *Learning for Life: The Achievements of 15-year-olds in Ireland on Mathematics, Reading Literacy and Science in PISA 2012*. Educational Research Centre.
- Scientific Software International. (2003). BILOG-MG (Version 3.0) [Computer software]. Author.
- Shany, M., & Biemiller, A. (2010). *Individual differences in reading comprehension gains from assisted reading practice: Pre existing conditions, vocabulary acquisition, and amounts of practice*. *Reading and Writing*, 23(9), 1071–1083. <https://doi.org/10.1007/s11145-009-9196-4>
- Shiel, G., Kavanagh, L., & Millar, D. (2014). *The 2014 National Assessments of English Reading and Mathematics: Vol. 1. Performance Report*. Educational Research Centre. http://www.erc.ie/wp-content/uploads/2016/11/NA_2014_Vol1_Final-updated.pdf
- Shiel, G., Kelleher, C., McKeown, C., & Denner, S. (2016). *Future Ready? The Performance of 15-year-olds in Ireland on Science, Reading Literacy and Mathematics in PISA 2015*. Educational Research Centre. https://www.erc.ie/wp-content/uploads/2016/12/PISA2015_FutureReady.pdf
- Shiel, G., Surgenor, P., Close, S., & Millar, D. (2006). *The 2004 National Assessments of Mathematics Achievement*. Educational Research Centre. https://www.erc.ie/documents/nama04_main_report_complete.pdf
- Verhoeven, L., & van Leeuwe, J. (2008). *Prediction of the development of reading comprehension: A longitudinal study*. *Applied Cognitive Psychology*, 22(3), 407–423. <https://doi.org/10.1002/acp.1414>
- What Works Clearinghouse. (2014). *Procedures and Standards Handbook. Version 3.0*. U.S. Department of Education, Institute of Education Sciences.
- Wolter, K. M. (2007). *Introduction to Variance Estimation* (2nd ed.). Springer.

Appendices

Appendix for Chapter 4

Second class English reading

Table A4.1. Mean percent correct scores for English reading components and process skills by year

| | 2009 | | 2014 | | 2021 | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Content area | | | | | | |
| Vocabulary | 63.1 | 0.78 | 69.7 | 0.67 | 63.2 | 0.68 |
| Comprehension | 62.8 | 0.80 | 67.5 | 0.71 | 62.3 | 0.75 |
| Process skill | | | | | | |
| Retrieve | 65.0 | 0.80 | 68.8 | 0.74 | 65.4 | 0.83 |
| Infer | 59.2 | 0.87 | 65.1 | 0.73 | 62.1 | 0.71 |
| Interpret & Integrate | 61.5 | 0.89 | 64.8 | 0.86 | 56.8 | 0.92 |
| Overall | 62.9 | 0.77 | 68.4 | 0.86 | 62.7 | 0.70 |

Table A4.2. Mean difficulty of the English reading trend items by year (2009, 2014 & 2021)

| Year | <i>% Correct</i> | <i>SD</i> |
|------|------------------|-----------|
| 2009 | 60.8 | 14.29 |
| 2014 | 67.0 | 14.63 |
| 2021 | 65.7 | 14.45 |

Note. % correct is based on the 94 trend items used in all three NAMER cycles

Table A4.3. Mean difficulty of the English reading trend items by year (2014 & 2021)

| Year | <i>% Correct</i> | <i>SD</i> |
|------|------------------|-----------|
| 2014 | 66.3 | 14.47 |
| 2021 | 65.2 | 14.41 |

Note. % correct is based on the 110 trend items used in the last two NAMER cycles

Table A4.4. Mean scale scores for English reading components and process skills by year

| | 2009 | | 2014 | | 2021 | | 2021 - 2014 | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | <i>95% CI</i> | <i>d</i> | |
| Content area | | | | | | | | | | | |
| Vocabulary | 250.0 | 1.72 | 264.9 | 1.55 | 262.6 | 1.50 | 2.3 | 2.54 | -2.7 | 7.3 | 0.05 |
| Comprehension | 250.0 | 1.70 | 262.5 | 1.61 | 259.2 | 1.66 | 3.3 | 2.31 | -1.3 | 7.9 | 0.07 |
| Process skill | | | | | | | | | | | |
| Retrieve | 250.0 | 1.63 | 260.5 | 1.54 | 257.8 | 1.62 | 2.7 | 2.24 | -1.7 | 7.1 | 0.06 |
| Infer | 250.0 | 1.71 | 263.2 | 1.48 | 260.9 | 1.58 | 2.3 | 2.16 | -1.9 | 6.6 | 0.05 |
| Interpret & Integrate | 250.0 | 1.61 | 261.1 | 1.40 | 260.1 | 1.68 | 1.0 | 2.19 | -3.3 | 5.4 | 0.02 |
| Overall | 250.0 | 1.77 | 264.0 | 1.65 | 260.8 | 1.65 | 3.2 | 2.33 | -1.4 | 7.8 | 0.07 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A4.5. English reading achievement for components and process skills at key percentile points by year

| | | Percentile | | | | | | | | | |
|-----------------------|------|------------|------|-------|------|-------|------|--------------|------|--------------|------|
| | | 10th | | 25th | | 50th | | 75th | | 90th | |
| | | SS | SE | SS | SE | SS | SE | SS | SE | SS | SE |
| Content area | | | | | | | | | | | |
| Vocabulary | 2009 | 186.6 | 1.68 | 208.9 | 1.98 | 247.3 | 2.47 | 284.7 | 2.54 | 319.1 | 2.13 |
| | 2014 | 197.4 | 2.20 | 229.0 | 2.08 | 264.6 | 1.90 | 299.4 | 1.87 | 331.7 | 2.09 |
| | 2021 | 199.2 | 2.30 | 228.0 | 2.21 | 262.3 | 2.59 | 297.1 | 2.40 | 326.7 | 2.52 |
| Comprehension | 2009 | 185.0 | 1.71 | 210.6 | 2.05 | 247.2 | 2.11 | 285.9 | 2.55 | 317.1 | 2.05 |
| | 2014 | 199.7 | 2.09 | 228.9 | 2.26 | 262.0 | 1.99 | 295.7 | 1.94 | 323.4 | 1.71 |
| | 2021 | 197.0 | 3.39 | 224.5 | 2.64 | 260.0 | 2.47 | 292.2 | 2.30 | 321.0 | 3.00 |
| Process skill | | | | | | | | | | | |
| Retrieve | 2009 | 182.8 | 1.65 | 209.8 | 2.42 | 248.0 | 2.22 | 288.7 | 2.06 | 321.2 | 2.81 |
| | 2014 | 197.8 | 2.29 | 227.1 | 2.03 | 260.8 | 1.84 | 296.4 | 1.83 | 320.4 | 2.71 |
| | 2021 | 196.5 | 2.98 | 225.4 | 3.01 | 258.5 | 2.30 | 290.0 | 2.02 | 317.4 | 2.62 |
| Infer | 2009 | 180.2 | 1.46 | 212.3 | 2.14 | 250.0 | 2.15 | 288.5 | 2.67 | 319.6 | 2.04 |
| | 2014 | 198.4 | 1.72 | 228.6 | 2.05 | 264.9 | 1.56 | 299.6 | 1.42 | 322.7 | 0.65 |
| | 2021 | 198.5 | 3.07 | 226.8 | 2.78 | 261.4 | 2.37 | 294.3 | 2.47 | 322.0 | 2.80 |
| Interpret & Integrate | 2009 | 174.8 | 2.80 | 207.5 | 2.58 | 251.7 | 2.20 | 295.3 | 1.39 | 313.5 | 1.11 |
| | 2014 | 201.0 | 0.40 | 231.8 | 3.47 | 260.8 | 2.36 | 296.6 | 0.13 | 313.0 | 0.54 |
| | 2021 | 198.9 | 3.90 | 227.7 | 2.64 | 260.9 | 2.30 | 292.0 | 2.33 | 320.7 | 2.76 |
| Overall | 2009 | 186.4 | 1.82 | 211.4 | 1.97 | 246.6 | 2.01 | 282.4 | 2.32 | 318.8 | 2.42 |
| | 2014 | 199.9 | 2.28 | 229.7 | 2.01 | 261.6 | 1.92 | 296.7 | 2.15 | 328.1 | 1.94 |
| | 2021 | 197.3 | 2.39 | 225.3 | 2.76 | 260.9 | 2.12 | 294.1 | 2.35 | 324.4 | 2.59 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Table A4.6. Percentages of pupils achieving at each proficiency level on the overall reading scale by year

| | 2009 | | 2014 | | 2021 | | 2021 - 2014 | | | |
|----------------------|------|------|------|------|------|------|-------------|------|--------|-----|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| Below Level 1 | 10.0 | 0.90 | 5.1 | 0.60 | 6.0 | 0.75 | 0.9 | 0.96 | -1.0 | 2.8 |
| Level 1 | 25.0 | 1.15 | 16.5 | 0.92 | 18.4 | 1.34 | 1.9 | 1.63 | -1.3 | 5.1 |
| Level 2 | 30.0 | 1.03 | 32.8 | 0.95 | 31.4 | 1.83 | 1.4 | 2.06 | -2.7 | 5.5 |
| Level 3 | 25.0 | 0.95 | 32.0 | 1.09 | 32.4 | 1.65 | 0.4 | 1.98 | -3.5 | 4.3 |
| Level 4 | 10.0 | 0.79 | 13.5 | 0.87 | 11.7 | 0.79 | 1.8 | 1.18 | -0.5 | 4.1 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A4.7. Percentages of pupils achieving at various proficiency levels on the overall reading scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------------|------|------|------|------|------|------|-------------|------|--------|-----|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| At or Below Level 1 | 35.0 | 1.59 | 21.7 | 1.22 | 24.4 | 1.58 | 2.7 | 1.99 | -1.2 | 6.7 |
| Level 2 | 30.0 | 1.03 | 32.8 | 0.95 | 31.4 | 1.83 | 1.4 | 2.06 | -2.7 | 5.5 |
| At or Above Level 3 | 35.0 | 1.37 | 45.5 | 1.49 | 44.1 | 1.76 | 1.4 | 2.31 | -3.2 | 5.9 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A4.8. Percentages of pupils achieving at various proficiency levels on the overall reading scale in 2009, 2014 and 2021, original and new targets

| | NAMER 2009 result | | Original 2020 target | NAMER 2014 result | | New 2020 target | NAMER 2021 result | |
|-------------------------------------|-------------------|------|----------------------|-------------------|------|-----------------|-------------------|------|
| | % | SE | % | % | SE | % | % | SE |
| Reading: at or above Level 3 | 35.0 | 1.37 | 40.0 | 45.5 | 1.49 | 50.0 | 44.1 | 1.76 |
| Reading: at or below Level 1 | 35.0 | 1.59 | 30.0 | 21.7 | 1.22 | 20.0 | 24.4 | 1.58 |

Note. 2021 percentages in bold are statistically significantly different from the new target percentages for 2020.

Sixth class Mathematics

Table A4.9. Mean percent correct scores for mathematics content areas and process skills by year

| | 2009 | | 2014 | | 2021 | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Content area | | | | | | |
| Number & Algebra | 57.6 | 1.05 | 63.4 | 0.89 | 63.1 | 0.73 |
| Shape & Space | 58.8 | 0.98 | 61.5 | 0.82 | 60.8 | 0.73 |
| Measures | 38.2 | 0.97 | 41.6 | 0.96 | 44.5 | 1.16 |
| Data | 63.5 | 0.99 | 62.6 | 0.87 | 61.4 | 0.95 |
| Process skill | | | | | | |
| Understand & Recall | 54.2 | 1.21 | 57.3 | 0.96 | 60.9 | 0.82 |
| Implement | 58.6 | 1.04 | 65.1 | 0.92 | 61.3 | 0.83 |
| Reason | 62.8 | 0.87 | 66.5 | 0.77 | 62.4 | 0.81 |
| Integrate & Connect | 58.5 | 1.18 | 65.2 | 1.03 | 66.3 | 0.74 |
| Apply & Problem Solve | 44.2 | 1.06 | 49.0 | 0.94 | 53.2 | 0.96 |
| Overall | 54.6 | 0.98 | 59.0 | 0.84 | 59.4 | 0.77 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Table A4.10. Mean difficulty of the mathematics trend items by year (2009, 2014 & 2021)

| Year | % Correct | <i>SD</i> |
|-------------|-----------|-----------|
| 2009 | 53.8 | 17.95 |
| 2014 | 58.8 | 17.96 |
| 2021 | 58.8 | 18.27 |

Note. % correct is based on the 127 trend items used in the last three NAMER cycles

Table A4.11. Mean difficulty of the mathematics trend items by year (2014 & 2021)

| Year | % Correct | <i>SD</i> |
|-------------|-----------|-----------|
| 2014 | 58.6 | 18.82 |
| 2021 | 58.5 | 19.10 |

Note. % correct is based on the 140 trend items used in the last two NAMER cycles

Table A4.12. Mean scale scores for mathematics content and process skills by year

| | 2009 | | 2014 | | 2021 | | 2021 - 2014 | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | <i>95% CI</i> | <i>d</i> | |
| Content area | | | | | | | | | | | |
| Number & Algebra | 250.0 | 2.34 | 260.8 | 2.07 | 259.0 | 1.70 | 1.8 | 2.68 | -3.5 | 7.1 | 0.04 |
| Shape & Space | 250.0 | 2.28 | 262.8 | 1.85 | 256.4 | 1.79 | 6.4 | 2.57 | 1.4 | 11.5 | 0.14 |
| Measures | 250.0 | 2.16 | 258.5 | 2.09 | 257.7 | 1.89 | 0.8 | 2.82 | -4.8 | 6.3 | 0.01 |
| Data | 250.0 | 1.99 | 259.0 | 1.77 | 257.3 | 1.84 | 1.7 | 2.55 | -3.4 | 6.7 | 0.03 |
| Process skill | | | | | | | | | | | |
| Understand & Recall | 250.0 | 2.44 | 260.2 | 1.90 | 258.0 | 1.60 | 2.2 | 2.49 | -2.7 | 7.1 | 0.05 |
| Implement | 250.0 | 2.31 | 262.8 | 2.12 | 259.9 | 1.76 | 2.9 | 2.76 | -2.5 | 8.4 | 0.06 |
| Reason | 250.0 | 2.07 | 261.6 | 1.82 | 260.4 | 2.09 | 1.2 | 2.77 | -4.3 | 6.7 | 0.02 |
| Integrate & Connect | 250.0 | 1.93 | 257.4 | 1.59 | 257.3 | 1.70 | 0.1 | 2.33 | -4.5 | 4.7 | 0.00 |
| Apply & Problem Solve | 250.0 | 2.35 | 259.2 | 2.10 | 258.3 | 1.77 | 0.9 | 2.75 | -4.5 | 6.3 | 0.02 |
| Overall | 250.0 | 2.35 | 261.7 | 2.07 | 260.5 | 1.82 | 1.2 | 2.75 | -4.3 | 6.6 | 0.02 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A4.13. Mathematics achievement for content areas and process skills at key percentile points by year

| Content area | | Percentile | | | | | | | | | |
|-----------------------|------|------------|------|--------------|------|--------------|------|--------------|------|--------------|------|
| | | 10th | | 25th | | 50th | | 75th | | 90th | |
| | | SS | SE | SS | SE | SS | SE | SS | SE | SS | SE |
| Number & Algebra | 2009 | 184.9 | 2.10 | 212.8 | 2.66 | 247.7 | 2.91 | 286.8 | 2.59 | 315.2 | 2.31 |
| | 2014 | 195.9 | 2.24 | 225.7 | 2.55 | 260.2 | 2.62 | 295.6 | 2.45 | 324.5 | 2.30 |
| | 2021 | 195.0 | 2.02 | 224.6 | 2.39 | 259.4 | 2.47 | 293.4 | 2.92 | 322.4 | 2.74 |
| Shape & Space | 2009 | 183.6 | 2.96 | 211.9 | 3.27 | 250.3 | 3.35 | 288.0 | 2.71 | 315.6 | 2.54 |
| | 2014 | 197.3 | 2.35 | 230.9 | 2.88 | 264.0 | 2.18 | 297.4 | 1.79 | 322.0 | 1.89 |
| | 2021 | 197.9 | 4.08 | 224.9 | 2.69 | 256.4 | 2.55 | 288.4 | 2.44 | 315.2 | 2.93 |
| Measures | 2009 | 184.8 | 1.25 | 210.4 | 2.05 | 245.9 | 2.89 | 286.3 | 2.97 | 319.6 | 2.63 |
| | 2014 | 191.9 | 1.81 | 218.0 | 1.83 | 254.9 | 2.71 | 295.9 | 2.54 | 326.7 | 3.06 |
| | 2021 | 192.5 | 2.83 | 222.4 | 2.89 | 257.6 | 2.94 | 293.7 | 2.81 | 323.9 | 3.57 |
| Data | 2009 | 180.7 | 3.57 | 212.9 | 3.04 | 251.7 | 2.82 | 290.0 | 2.51 | 311.6 | 2.22 |
| | 2014 | 192.3 | 1.30 | 226.3 | 2.21 | 261.7 | 3.00 | 296.1 | 1.12 | 325.2 | 1.44 |
| | 2021 | 193.4 | 2.89 | 222.7 | 3.06 | 257.6 | 2.58 | 291.9 | 2.70 | 321.9 | 3.31 |
| Process skill | | | | | | | | | | | |
| Understand & Recall | 2009 | 175.9 | 3.70 | 216.8 | 3.74 | 257.3 | 1.54 | 297.3 | 3.06 | 306.0 | 1.70 |
| | 2014 | 195.1 | 2.93 | 227.6 | 2.74 | 262.4 | 2.44 | 293.9 | 1.79 | 321.3 | 1.86 |
| | 2021 | 196.3 | 2.36 | 225.9 | 2.56 | 258.8 | 2.52 | 290.9 | 2.92 | 318.2 | 2.94 |
| Implement | 2009 | 180.5 | 4.23 | 215.2 | 3.60 | 251.8 | 2.93 | 286.6 | 2.36 | 312.2 | 1.69 |
| | 2014 | 196.2 | 2.99 | 226.9 | 2.29 | 264.8 | 3.08 | 300.1 | 2.22 | 326.6 | 2.38 |
| | 2021 | 195.0 | 2.02 | 224.9 | 2.59 | 261.3 | 2.49 | 294.7 | 3.24 | 323.4 | 3.40 |
| Reason | 2009 | 184.8 | 2.27 | 213.1 | 2.54 | 248.3 | 2.82 | 284.7 | 3.86 | 319.2 | 3.08 |
| | 2014 | 195.0 | 2.13 | 227.9 | 2.35 | 263.1 | 2.77 | 296.9 | 2.24 | 325.0 | 2.02 |
| | 2021 | 193.6 | 3.06 | 225.9 | 2.62 | 261.1 | 2.64 | 296.1 | 2.84 | 325.4 | 3.90 |
| Integrate & Connect | 2009 | 181.8 | 2.92 | 217.0 | 3.13 | 250.6 | 3.84 | 285.8 | 5.16 | 316.3 | 3.11 |
| | 2014 | 196.6 | 5.25 | 216.2 | 0.27 | 261.9 | 1.27 | 298.7 | 0.90 | 304.1 | 0.74 |
| | 2021 | 193.4 | 3.46 | 224.5 | 2.53 | 257.7 | 2.45 | 290.4 | 3.12 | 319.2 | 4.62 |
| Apply & Problem Solve | 2009 | 182.4 | 3.12 | 211.0 | 3.18 | 249.5 | 2.78 | 284.8 | 2.90 | 317.5 | 3.86 |
| | 2014 | 192.8 | 1.99 | 222.6 | 2.34 | 257.3 | 2.82 | 294.4 | 2.34 | 325.4 | 2.65 |
| | 2021 | 193.1 | 2.92 | 223.5 | 2.83 | 258.3 | 2.53 | 293.3 | 2.73 | 322.3 | 3.24 |
| Overall | 2009 | 183.4 | 1.94 | 213.8 | 2.93 | 248.7 | 2.71 | 285.6 | 2.35 | 314.5 | 2.31 |
| | 2014 | 198.0 | 2.25 | 227.3 | 2.18 | 261.3 | 2.89 | 296.6 | 2.22 | 326 | 2.56 |
| | 2021 | 195.5 | 2.71 | 225.5 | 2.87 | 260.9 | 2.56 | 295.5 | 2.37 | 325.0 | 2.87 |

Note. 2021 scores in bold are statistically significantly different from the corresponding 2014 score.

Table A4.14. Percentage of pupils achieving at each proficiency level on the overall mathematics scale by year

| | 2009 | | 2014 | | 2021 | | 2021 - 2014 | | | |
|----------------------|------|------|------|------|------|------|-------------|------|--------|-----|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| Below Level 1 | 10.0 | 0.97 | 5.6 | 0.65 | 6.2 | 0.63 | 0.6 | 0.90 | -1.2 | 2.4 |
| Level 1 | 25.0 | 1.70 | 20.9 | 1.36 | 21.1 | 1.58 | 0.2 | 2.09 | -3.9 | 4.3 |
| Level 2 | 30.0 | 1.17 | 31.4 | 1.31 | 31.3 | 1.31 | 0.1 | 1.85 | -3.6 | 3.8 |
| Level 3 | 25.0 | 1.46 | 27.2 | 1.18 | 27.3 | 1.37 | 0.1 | 1.81 | -3.4 | 3.7 |
| Level 4 | 10.0 | 0.85 | 14.9 | 1.14 | 14.1 | 1.28 | 0.8 | 1.72 | -2.6 | 4.2 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A4.15. Percentages of pupils achieving at various proficiency levels on the overall mathematics scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------------|------|------|------|------|------|------|-------------|------|--------|-----|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| At or Below Level 1 | 35.0 | 2.25 | 26.5 | 1.61 | 27.3 | 1.65 | 0.8 | 2.31 | -3.8 | 5.3 |
| Level 2 | 30.0 | 1.17 | 31.4 | 1.31 | 31.3 | 1.31 | 0.1 | 1.85 | -3.6 | 3.8 |
| At or Above Level 3 | 35.0 | 1.87 | 42.1 | 1.97 | 41.4 | 1.54 | 0.7 | 2.50 | -4.2 | 5.6 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A4.16. Percentages of Sixth class pupils achieving at various proficiency levels on the overall mathematics scale in 2009, 2014 and 2021, 2011 and 2014 targets.

| | NAMER 2009 result | | Original 2020 target | NAMER 2014 result | | New 2020 target | NAMER 2021 result | |
|---|-------------------|------|----------------------|-------------------|------|-----------------|-------------------|------|
| | % | SE | % | % | SE | % | % | SE |
| Mathematics: at or above Level 3 | 35.0 | 1.87 | 40.0 | 42.1 | 1.97 | 50.0 | 41.4 | 1.54 |
| Mathematics: at or below Level 1 | 35.0 | 2.25 | 30.0 | 26.5 | 1.61 | 20.0 | 27.3 | 1.65 |

Note. 2021 percentages in bold are statistically significantly different from the new target percentages for 2020.

Appendix for Chapter 5

Second class English reading

Table A5.1. Mean percent correct scores for English reading components and process skills by gender, NAMER 2021

| | Boys | | Girls | | Other | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Content Area | | | | | | |
| Vocabulary | 62.3 | 0.82 | 64.1 | 0.89 | 62.1 | 6.36 |
| Comprehension | 59.9 | 0.80 | 64.9 | 1.33 | 62.9 | 5.69 |
| Process Skill | | | | | | |
| Retrieve | 63.2 | 0.91 | 67.7 | 1.34 | 63.6 | 6.41 |
| Infer | 59.4 | 0.79 | 65.0 | 1.25 | 64.3 | 5.68 |
| Interpret & Integrate | 54.2 | 0.88 | 59.5 | 1.57 | 56.1 | 6.28 |
| Overall | 60.8 | 0.76 | 64.6 | 1.15 | 62.5 | 5.89 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16.

Table A5.2. Mean scale scores for English reading components and process skills by gender, NAMER 2021

| | Boys | | Girls | | Other | | Boys - Girls | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|--------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | <i>95% CI</i> | <i>d</i> | |
| Content area | | | | | | | | | | | |
| Vocabulary | 260.8 | 1.98 | 264.6 | 2.22 | 261.2 | 16.64 | 3.7 | 2.86 | -1.9 | 9.4 | 0.08 |
| Comprehension | 254.8 | 2.01 | 263.9 | 2.95 | 260.4 | 13.47 | 9.1 | 3.61 | 2.0 | 16.3 | 0.19 |
| Process skill | | | | | | | | | | | |
| Retrieve | 254.6 | 2.26 | 261.2 | 2.69 | 256.7 | 14.03 | 6.6 | 3.64 | -0.6 | 13.8 | 0.14 |
| Infer | 257.1 | 2.19 | 264.9 | 2.50 | 262.8 | 14.21 | 7.7 | 3.37 | 1.1 | 14.4 | 0.16 |
| Interpret & Integrate | 257.2 | 2.23 | 263.1 | 3.03 | 263.2 | 14.28 | 5.9 | 3.97 | -1.9 | 13.7 | 0.13 |
| Overall | 256.8 | 2.00 | 265.1 | 2.78 | 261.1 | 14.55 | 8.3 | 3.34 | 1.7 | 14.9 | 0.17 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Bold font indicates a statistically significant difference between boys and girls.

Table A5.3. Percentages of pupils achieving at each proficiency level on the overall reading scale by gender, NAMER 2021

| | Boys | | Girls | | Other | | Boys - Girls | | | |
|----------------------|------|------|-------|------|-------|-------|--------------|-----|--------|------|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| Below Level 1 | 6.6 | 0.95 | 5.3 | 1.18 | 10.2 | 7.11 | 1.3 | 1.5 | -1.7 | 4.3 |
| Level 1 | 21.7 | 2.12 | 15.0 | 1.89 | 17.8 | 10.38 | 6.7 | 2.9 | 0.9 | 12.4 |
| Level 2 | 30.5 | 2.99 | 32.5 | 2.19 | 20.6 | 10.7 | 2.0 | 3.8 | -5.5 | 9.5 |
| Level 3 | 30.8 | 2.36 | 34.1 | 2.42 | 42.1 | 15.08 | 3.3 | 3.5 | -3.6 | 10.1 |
| Level 4 | 10.4 | 1.24 | 13.1 | 1.41 | 9.3 | 11.08 | 2.7 | 2.1 | -1.4 | 6.9 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Bold font indicates a statistically significant difference between boys and girls.

Table A5.4. Percentages of pupils achieving at different proficiency levels on the overall reading scale by gender, NAMER 2021

| | Boys | | Girls | | Other | | Boys - Girls | | | |
|----------------------------|------|------|-------|------|-------|-------|--------------|------|--------|------|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| At or Below Level 1 | 28.3 | 2.12 | 20.3 | 2.41 | 28.0 | 11.13 | 8.0 | 3.09 | 1.9 | 14.1 |
| Level 2 | 30.5 | 2.99 | 32.5 | 2.19 | 20.6 | 10.70 | 2.0 | 3.79 | -5.5 | 9.5 |
| At or Above Level 3 | 41.2 | 2.41 | 47.2 | 2.73 | 51.4 | 14.11 | 6.0 | 3.71 | -1.3 | 13.3 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Bold font indicates a statistically significant difference between boys and girls.

Table A5.5. Mean percent correct scores for English Reading components and process skills by gender and year

| Content area | 2009 | | | | 2014 | | | | 2021 | | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|
| | Boys | | Girls | | Boys | | Girls | | Boys | | Girls | | Other | |
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Vocabulary | 61.2 | 0.94 | 65.0 | 0.99 | 69.2 | 0.83 | 70.2 | 0.88 | 62.3 | 0.82 | 64.1 | 0.89 | 62.1 | 6.36 |
| Comprehension | 59.2 | 1.13 | 66.5 | 0.97 | 65.5 | 0.87 | 69.5 | 0.80 | 59.9 | 0.80 | 64.9 | 1.33 | 62.9 | 5.69 |
| Process Skill | | | | | | | | | | | | | | |
| Retrieve | 61.3 | 1.19 | 68.8 | 0.96 | 67.0 | 0.92 | 70.7 | 0.81 | 63.2 | 0.91 | 67.7 | 1.34 | 63.6 | 6.41 |
| Infer | 56.3 | 1.15 | 62.2 | 1.07 | 62.9 | 0.85 | 67.2 | 0.86 | 59.4 | 0.79 | 65.0 | 1.25 | 64.3 | 5.68 |
| Interpret & Integrate | 57.4 | 1.16 | 65.8 | 1.09 | 62.0 | 1.06 | 67.6 | 1.08 | 54.2 | 0.88 | 59.5 | 1.57 | 56.1 | 6.28 |
| Overall | 60.0 | 1.02 | 65.9 | 0.95 | 66.9 | 0.83 | 69.8 | 0.80 | 60.8 | 0.76 | 64.6 | 1.15 | 62.5 | 5.89 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Bold font indicates a statistically significant difference between boys and girls.

Table A5.6. Boys' mean scale scores for English Reading components and process skills by year

| Content area | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | 95% <i>CI</i> | <i>d</i> | |
| Vocabulary | 245.8 | 2.13 | 263.3 | 1.96 | 260.8 | 1.98 | 2.5 | 2.8 | -3.0 | 8.0 | 0.05 |
| Comprehension | 242.6 | 2.39 | 258.1 | 1.95 | 254.8 | 2.01 | 3.3 | 2.8 | -2.2 | 8.9 | 0.07 |
| Process skill | | | | | | | | | | | |
| Retrieve | 242.8 | 2.45 | 256.8 | 1.89 | 254.6 | 2.26 | 2.2 | 2.9 | -3.6 | 8.0 | 0.05 |
| Infer | 244.6 | 2.29 | 258.8 | 1.75 | 257.1 | 2.19 | 1.7 | 2.8 | -3.9 | 7.2 | 0.04 |
| Interpret & Integrate | 243.5 | 2.07 | 257.2 | 1.67 | 257.2 | 2.23 | 0.0 | 2.8 | -5.5 | 5.5 | 0.0 |
| Overall | 243.4 | 2.31 | 260.4 | 2.01 | 256.8 | 2.00 | 3.6 | 2.8 | -2.0 | 9.2 | 0.07 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.7. Girls' mean scale scores for English Reading components and process skills by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | <i>95% CI</i> | <i>d</i> | |
| Content area | | | | | | | | | | | |
| Vocabulary | 254.4 | 2.21 | 266.6 | 2.05 | 264.6 | 2.22 | 2.1 | 3.0 | -3.9 | 8.0 | 0.04 |
| Comprehension | 257.7 | 2.16 | 266.8 | 1.81 | 263.9 | 2.95 | 2.9 | 3.5 | -3.9 | 9.7 | 0.06 |
| Process skill | | | | | | | | | | | |
| Retrieve | 257.6 | 2.12 | 264.1 | 1.77 | 261.2 | 2.69 | 2.9 | 3.2 | -3.4 | 9.3 | 0.07 |
| Infer | 255.7 | 2.04 | 267.6 | 1.68 | 264.9 | 2.50 | 2.7 | 3.0 | -3.2 | 8.7 | 0.06 |
| Interpret & Integrate | 256.8 | 1.77 | 265.0 | 1.71 | 263.1 | 3.03 | 1.9 | 3.5 | -4.9 | 8.8 | 0.04 |
| Overall | 256.9 | 2.29 | 267.5 | 1.98 | 265.1 | 2.78 | 2.4 | 3.4 | -4.3 | 9.2 | 0.05 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.8. Percentages of boys performing at each proficiency level on the overall reading scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|-----|
| | <i>%</i> | <i>SE</i> | <i>%</i> | <i>SE</i> | <i>%</i> | <i>SE</i> | <i>PD</i> | <i>SED</i> | <i>95% CI</i> | |
| Below Level 1 | 12.5 | 1.30 | 6.5 | 0.84 | 6.6 | 0.95 | 0.1 | 1.27 | -2.4 | 2.6 |
| Level 1 | 27.8 | 1.64 | 16.9 | 1.24 | 21.7 | 2.12 | 4.8 | 2.45 | -0.1 | 9.6 |
| Level 2 | 29.5 | 1.73 | 33.8 | 1.51 | 30.5 | 2.99 | 3.3 | 3.35 | -3.3 | 9.9 |
| Level 3 | 22.5 | 1.71 | 31.3 | 1.32 | 30.8 | 2.36 | 0.5 | 2.70 | -4.9 | 5.8 |
| Level 4 | 7.7 | 0.88 | 11.6 | 1.18 | 10.4 | 1.24 | 1.2 | 1.71 | -2.2 | 4.6 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.9. Percentages of girls performing at each proficiency level on the overall reading scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|-----|
| | <i>%</i> | <i>SE</i> | <i>%</i> | <i>SE</i> | <i>%</i> | <i>SE</i> | <i>PD</i> | <i>SED</i> | <i>95% CI</i> | |
| Below Level 1 | 6.9 | 0.90 | 3.8 | 0.56 | 5.3 | 1.18 | 1.5 | 1.31 | -1.1 | 4.1 |
| Level 1 | 22.6 | 1.52 | 16.2 | 1.24 | 15.0 | 1.89 | 1.2 | 2.26 | -3.2 | 5.7 |
| Level 2 | 30.5 | 1.92 | 31.9 | 1.07 | 32.5 | 2.19 | 0.6 | 2.44 | -4.3 | 5.4 |
| Level 3 | 27.7 | 1.68 | 32.7 | 1.43 | 34.1 | 2.42 | 1.4 | 2.81 | -4.2 | 6.9 |
| Level 4 | 12.2 | 1.14 | 15.4 | 1.13 | 13.1 | 1.41 | 2.3 | 1.80 | -1.3 | 5.8 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.10. Percentages of boys performing at various proficiency levels on the overall reading scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------------|------|------|------|------|-------------|------|-------------|------|--------|------|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| At or Below Level 1 | 40.3 | 2.10 | 23.3 | 1.57 | 28.3 | 2.12 | 5.0 | 2.64 | -0.2 | 10.2 |
| Level 2 | 29.5 | 1.73 | 33.8 | 1.51 | 30.5 | 2.99 | 3.3 | 3.35 | -3.3 | 9.9 |
| At or Above Level 3 | 30.2 | 2.12 | 42.9 | 1.86 | 41.2 | 2.41 | 1.6 | 3.05 | -4.4 | 7.7 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.11. Percentages of girls performing at various proficiency levels on the overall reading scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------------|------|------|------|------|------|------|-------------|------|--------|-----|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| At or Below Level 1 | 29.5 | 1.83 | 20.0 | 1.49 | 20.3 | 2.41 | 0.3 | 2.84 | -5.3 | 5.9 |
| Level 2 | 30.5 | 1.98 | 31.9 | 1.07 | 32.5 | 2.19 | 0.6 | 2.44 | -4.2 | 5.4 |
| At or Above Level 3 | 40.0 | 2.17 | 48.1 | 1.78 | 47.2 | 2.73 | 0.9 | 3.26 | -5.5 | 7.3 |

Note. 2021 percentages in bold are statistically significantly different from the corresponding 2014 percentage.

Sixth class Mathematics

Table A5.12. Mean percent correct scores for mathematics content areas and process skills by gender, NAMER 2021

| | Boys | | Girls | | Other | |
|-----------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Content area | | | | | | |
| Number & Algebra | 65.0 | 0.91 | 61.4 | 0.93 | 59.1 | 3.23 |
| Shape & Space | 61.1 | 0.86 | 60.5 | 0.96 | 62.9 | 3.26 |
| Measures | 47.4 | 1.21 | 41.8 | 1.35 | 39.6 | 3.76 |
| Data | 63.0 | 1.45 | 59.9 | 1.00 | 62.5 | 5.26 |
| Process Skill | | | | | | |
| Understand & Recall | 63.2 | 0.92 | 58.7 | 1.17 | 60.4 | 3.96 |
| Implement | 62.0 | 1.13 | 60.8 | 0.94 | 58.0 | 3.24 |
| Reason | 64.1 | 1.06 | 60.8 | 0.93 | 59.7 | 3.67 |
| Integrate & Connect | 67.3 | 0.96 | 65.3 | 1.28 | 68.3 | 4.76 |
| Apply & Problem Solve | 55.7 | 0.97 | 51.0 | 1.29 | 49.8 | 3.75 |
| Overall | 61.2 | 0.91 | 57.8 | 0.94 | 56.7 | 3.07 |

Note. N(Boys) = 2618, N(Girls) = 2697, N(Other) = 65, except Integrate & Connect N(Boys) = 1958, N(Girls) = 1994, N(Other) = 50.

Table A5.13. Mean scale scores for mathematics content areas and process skills by gender, NAMER 2021

| | Boys | | Girls | | Other | | Boys - Girls | | | | |
|-----------------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | <i>95% CI</i> | <i>d</i> | |
| Content area | | | | | | | | | | | |
| Number | 263.2 | 2.13 | 255.2 | 2.17 | 251.4 | 7.27 | 8.0 | 2.65 | 2.8 | 13.3 | 0.16 |
| Shape & Space | 257.3 | 1.85 | 255.5 | 2.37 | 255.4 | 7.99 | 1.9 | 2.33 | -2.7 | 6.5 | 0.04 |
| Measures | 262.3 | 1.97 | 253.4 | 2.51 | 253.7 | 8.84 | 8.8 | 2.69 | 3.5 | 14.1 | 0.17 |
| Data | 259.0 | 2.37 | 255.7 | 2.44 | 258.8 | 10.08 | 3.3 | 3.15 | -2.9 | 9.6 | 0.07 |
| Process skill | | | | | | | | | | | |
| Understand & Recall | 261.5 | 2.00 | 254.5 | 2.44 | 260.3 | 8.36 | 7.0 | 3.11 | 0.9 | 13.1 | 0.15 |
| Implement | 261.3 | 2.34 | 258.6 | 2.29 | 255.7 | 8.44 | 2.7 | 2.99 | -3.2 | 8.6 | 0.05 |
| Reason | 263.9 | 2.50 | 257.0 | 2.52 | 255.9 | 8.09 | 6.9 | 2.85 | 1.3 | 12.6 | 0.14 |
| Integrate & Connect | 259.0 | 2.35 | 255.6 | 2.38 | 257.7 | 10.96 | 3.5 | 3.17 | -2.8 | 9.7 | 0.07 |
| Apply & Problem Solve | 262.6 | 1.99 | 254.2 | 2.45 | 253.0 | 7.68 | 8.4 | 2.85 | 2.8 | 14.1 | 0.17 |
| Overall | 264.6 | 2.20 | 256.8 | 2.21 | 253.9 | 7.20 | 7.8 | 2.59 | 2.7 | 12.9 | 0.16 |

Note. N(Boys) = 2618, N(Girls) = 2697, N(Other) = 65, except Integrate & Connect N(Boys) = 1958, N(Girls) = 1994, N(Other) = 50. Bold font indicates a statistically significant difference between boys and girls.

Table A5.14. Percentages of pupils achieving at each proficiency level on the overall mathematics scale by gender, NAMER 2021

| | Boys | | Girls | | Other | | Boys - Girls | | | | |
|----------------------|----------|-----------|----------|-----------|----------|-----------|--------------|------------|---------------|------|--|
| | <i>%</i> | <i>SE</i> | <i>%</i> | <i>SE</i> | <i>%</i> | <i>SE</i> | <i>PD</i> | <i>SED</i> | <i>95% CI</i> | | |
| Below Level 1 | 6.0 | 0.91 | 6.3 | 0.93 | 5.5 | 3.68 | 0.3 | 1.31 | -2.6 | 3.3 | |
| Level 1 | 17.1 | 1.58 | 24.9 | 2.14 | 26.4 | 7.28 | 7.7 | 2.30 | 2.5 | 12.9 | |
| Level 2 | 31.7 | 1.76 | 30.9 | 1.95 | 30.2 | 8.19 | 0.8 | 2.62 | -5.1 | 6.7 | |
| Level 3 | 29.6 | 2.10 | 25.2 | 1.84 | 28.1 | 7.52 | 4.4 | 2.77 | -1.8 | 10.7 | |
| Level 4 | 15.6 | 1.77 | 12.8 | 1.71 | 9.9 | 5.19 | 2.8 | 2.35 | -2.5 | 8.1 | |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Bold font indicates a statistically significant difference between boys and girls.

Table A5.15. Percentages of pupils achieving at different proficiency levels on the overall mathematics scale by gender, NAMER 2021

| | Boys | | Girls | | Other | | Boys - Girls | | | |
|---------------------|------|------|-------|------|-------|------|--------------|------|--------|------|
| | % | SE | % | SE | % | SE | PD | SED | 95% CI | |
| At or Below Level 1 | 23.1 | 1.68 | 31.2 | 2.07 | 31.9 | 6.70 | 8.1 | 2.10 | 3.9 | 12.2 |
| Level 2 | 31.7 | 1.76 | 30.9 | 1.95 | 30.2 | 8.19 | 0.8 | 2.62 | -4.3 | 6.0 |
| At or Above Level 3 | 45.2 | 2.26 | 37.9 | 1.97 | 38.0 | 8.15 | 7.3 | 2.92 | 1.5 | 13.0 |

Note. N(Boys) = 2560, N(Girls) = 2452, N(Other) = 16. Bold font indicates a statistically significant difference between boys and girls.

Table A5.16. Mean percent correct scores for mathematics content areas and process skills by gender and year

| | 2009 | | | | 2014 | | | | 2021 | | | | Other | |
|-----------------------|------|------|-------|------|------|------|-------|------|------|------|-------|------|-------|------|
| | Boys | | Girls | | Boys | | Girls | | Boys | | Girls | | | |
| | M | SE | M | SE | M | SE | M | SE | M | SE | M | SE | M | SE |
| Content area | | | | | | | | | | | | | | |
| Number & Algebra | 58.6 | 1.12 | 56.5 | 1.50 | 63.9 | 1.09 | 63.0 | 1.12 | 65.0 | 0.91 | 61.4 | 0.93 | 59.1 | 3.23 |
| Shape & Space | 59.2 | 1.01 | 58.4 | 1.33 | 61.6 | 0.95 | 61.5 | 1.04 | 61.1 | 0.86 | 60.5 | 0.96 | 62.9 | 3.26 |
| Measures | 40.3 | 1.16 | 36.0 | 1.22 | 43.1 | 1.20 | 40.3 | 1.15 | 47.4 | 1.21 | 41.8 | 1.35 | 39.6 | 3.76 |
| Data | 64.0 | 1.13 | 63.1 | 1.58 | 62.9 | 0.96 | 62.3 | 1.14 | 63.0 | 1.45 | 59.9 | 1.00 | 62.5 | 5.26 |
| Process Skill | | | | | | | | | | | | | | |
| Understand & Recall | 54.9 | 1.23 | 53.5 | 1.60 | 57.4 | 1.21 | 57.2 | 1.24 | 63.2 | 0.92 | 58.7 | 1.17 | 60.4 | 3.96 |
| Implement | 59.1 | 1.21 | 58.1 | 1.39 | 64.7 | 1.13 | 65.4 | 1.11 | 62.0 | 1.13 | 60.8 | 0.94 | 58.0 | 3.24 |
| Reason | 63.3 | 0.96 | 62.4 | 1.29 | 66.4 | 0.96 | 66.6 | 0.99 | 64.1 | 1.06 | 60.8 | 0.93 | 59.7 | 3.67 |
| Integrate & Connect | 59.7 | 1.24 | 57.1 | 1.87 | 65.7 | 1.23 | 64.9 | 1.35 | 67.3 | 0.96 | 65.3 | 1.28 | 68.3 | 4.76 |
| Apply & Problem Solve | 46.4 | 1.18 | 42.0 | 1.45 | 50.9 | 1.08 | 47.3 | 1.14 | 55.7 | 0.97 | 51.0 | 1.29 | 49.8 | 3.75 |
| Overall | 55.7 | 1.05 | 53.5 | 1.37 | 59.6 | 1.00 | 58.5 | 1.06 | 61.2 | 0.91 | 57.8 | 0.94 | 56.7 | 3.07 |

Note. 2021 percentages in bold are statistically significantly different from the corresponding 2014 percentage.

Table A5.17. Girls' mean scale scores for mathematics content area and process skills by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | <i>95% CI</i> | <i>d</i> | |
| Content area | | | | | | | | | | | |
| Number & Algebra | 247.2 | 3.28 | 259.2 | 2.59 | 255.2 | 2.17 | 4.0 | 3.38 | -2.6 | 10.7 | 0.08 |
| Shape & Space | 249.1 | 3.10 | 263.2 | 2.32 | 255.4 | 2.37 | 7.8 | 3.32 | 1.2 | 14.3 | 0.17 |
| Measures | 245.5 | 2.78 | 255.4 | 2.52 | 253.4 | 2.51 | 2.0 | 3.56 | -5.1 | 9.0 | 0.04 |
| Data | 248.6 | 3.18 | 258.4 | 2.30 | 255.7 | 2.44 | 2.7 | 3.35 | -3.9 | 9.3 | 0.05 |
| Process Skill | | | | | | | | | | | |
| Understand & Recall | 248.5 | 2.18 | 260.1 | 2.47 | 254.5 | 2.44 | 5.6 | 3.47 | -1.3 | 12.4 | 0.12 |
| Implement | 248.9 | 3.11 | 263.2 | 2.56 | 258.6 | 2.29 | 4.6 | 3.44 | -2.2 | 11.4 | 0.09 |
| Reason | 248.2 | 3.10 | 261.4 | 2.46 | 257.0 | 2.52 | 4.4 | 3.52 | -2.5 | 11.4 | 0.09 |
| Integrate & Connect | 247.6 | 2.99 | 257.0 | 2.21 | 255.6 | 2.38 | 1.4 | 3.25 | -5.0 | 7.8 | 0.03 |
| Apply & Problem Solve | 245.4 | 3.27 | 255.6 | 2.54 | 254.2 | 2.45 | 1.4 | 3.53 | -5.6 | 8.3 | 0.03 |
| Overall | 247.0 | 3.27 | 260.0 | 2.56 | 256.8 | 2.21 | 3.2 | 3.38 | -3.5 | 9.9 | 0.07 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.18. Boys' mean scale scores for mathematics content area and process skills by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | | |
|-----------------------|----------|-----------|----------|-----------|----------|-----------|-------------|------------|---------------|----------|------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>MD</i> | <i>SED</i> | <i>95% CI</i> | <i>d</i> | |
| Content area | | | | | | | | | | | |
| Number & Algebra | 252.6 | 2.56 | 262.6 | 2.56 | 263.2 | 2.13 | 0.6 | 3.33 | -6.0 | 7.2 | 0.01 |
| Shape & Space | 250.8 | 2.39 | 262.4 | 2.19 | 257.3 | 1.85 | 5.1 | 2.87 | -0.6 | 10.7 | 0.11 |
| Measures | 254.2 | 2.46 | 262.1 | 2.6 | 262.3 | 1.97 | 0.2 | 3.26 | -6.3 | 6.6 | 0.00 |
| Data | 251.3 | 2.22 | 259.7 | 2.08 | 259.0 | 2.37 | 0.7 | 3.16 | -5.5 | 6.9 | 0.01 |
| Process Skill | | | | | | | | | | | |
| Understand & Recall | 251.5 | 2.52 | 260.4 | 2.36 | 261.5 | 2.00 | 1.1 | 3.1 | -5.0 | 7.2 | 0.02 |
| Implement | 251.1 | 2.71 | 262.3 | 2.57 | 261.3 | 2.34 | 1.0 | 3.5 | -5.8 | 7.9 | 0.02 |
| Reason | 251.6 | 2.23 | 261.7 | 2.27 | 263.9 | 2.50 | 2.2 | 3.4 | -4.5 | 8.9 | 0.04 |
| Integrate & Connect | 252.3 | 1.86 | 257.9 | 1.86 | 259.0 | 2.35 | 1.1 | 3.0 | -4.8 | 7.0 | 0.02 |
| Apply & Problem Solve | 254.3 | 2.55 | 263.4 | 2.51 | 262.6 | 1.99 | 0.7 | 3.2 | -5.6 | 7.1 | 0.01 |
| Overall | 252.8 | 2.55 | 263.7 | 2.54 | 264.6 | 2.20 | 0.9 | 3.36 | -5.7 | 7.5 | 0.02 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.19. Percentages of boys performing at each proficiency level on the overall mathematics scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------|------|-----------|------|-----------|------|-----------|-------------|------------|---------------|-----|
| | % | <i>SE</i> | % | <i>SE</i> | % | <i>SE</i> | <i>PD</i> | <i>SED</i> | <i>95% CI</i> | |
| Below Level 1 | 9.4 | 1.16 | 6.7 | 0.96 | 6.0 | 0.91 | 0.7 | 1.32 | -1.9 | 3.3 |
| Level 1 | 23.2 | 1.85 | 19.1 | 1.64 | 17.1 | 1.58 | 2.0 | 2.28 | -2.5 | 6.5 |
| Level 2 | 29.3 | 1.26 | 30.7 | 1.53 | 31.7 | 1.76 | 1.1 | 2.33 | -3.5 | 5.7 |
| Level 3 | 26.9 | 1.75 | 26.6 | 1.48 | 29.6 | 2.10 | 3.0 | 2.56 | -2.1 | 8.1 |
| Level 4 | 11.1 | 1.06 | 17.0 | 1.63 | 15.6 | 1.77 | 1.4 | 2.41 | -3.4 | 6.2 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.20. Percentages of girls performing at each proficiency level on the overall mathematics scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------|------|-----------|------|-----------|------|-----------|-------------|------------|---------------|-----|
| | % | <i>SE</i> | % | <i>SE</i> | % | <i>SE</i> | <i>PD</i> | <i>SED</i> | <i>95% CI</i> | |
| Below Level 1 | 10.7 | 1.67 | 4.6 | 0.66 | 6.3 | 0.93 | 1.7 | 1.14 | -0.5 | 4.0 |
| Level 1 | 26.9 | 2.00 | 22.5 | 1.67 | 24.9 | 2.14 | 2.4 | 2.71 | -3.0 | 7.7 |
| Level 2 | 30.6 | 1.79 | 32.0 | 1.76 | 30.9 | 1.95 | 1.1 | 2.63 | -4.1 | 6.3 |
| Level 3 | 23.0 | 1.91 | 27.7 | 1.66 | 25.2 | 1.84 | 2.5 | 2.48 | -2.4 | 7.4 |
| Level 4 | 8.8 | 1.11 | 13.2 | 1.22 | 12.8 | 1.71 | 0.4 | 2.10 | -3.7 | 4.6 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.21. Percentages of boys performing at various proficiency levels on the overall mathematics scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------------|------|-----------|------|-----------|------|-----------|-------------|------------|---------------|-----|
| | % | <i>SE</i> | % | <i>SE</i> | % | <i>SE</i> | <i>PD</i> | <i>SED</i> | <i>95% CI</i> | |
| At or Below Level 1 | 32.6 | 2.06 | 25.8 | 2.04 | 23.1 | 1.68 | 2.7 | 2.65 | -2.5 | 7.9 |
| Level 2 | 29.3 | 1.26 | 30.7 | 1.53 | 31.7 | 1.76 | 1.1 | 2.33 | -3.5 | 5.7 |
| At or Above Level 3 | 38.0 | 2.27 | 43.6 | 2.24 | 45.2 | 2.26 | 1.6 | 3.18 | -4.7 | 7.9 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

Table A5.22. Percentages of girls performing at various proficiency levels on the overall mathematics scale by year

| | 2009 | | 2014 | | 2021 | | 2014 - 2021 | | | |
|----------------------------|------|-----------|------|-----------|------|-----------|-------------|------------|---------------|-----|
| | % | <i>SE</i> | % | <i>SE</i> | % | <i>SE</i> | <i>PD</i> | <i>SED</i> | <i>95% CI</i> | |
| At or Below Level 1 | 37.6 | 3.28 | 27.1 | 1.97 | 31.2 | 2.07 | 4.1 | 2.86 | -1.6 | 9.7 |
| Level 2 | 30.6 | 1.79 | 32.0 | 1.76 | 30.9 | 1.95 | 1.1 | 2.63 | -4.1 | 6.3 |
| At or Above Level 3 | 31.8 | 2.43 | 40.9 | 2.39 | 37.9 | 1.97 | 3.0 | 3.09 | -3.1 | 9.1 |

Note. Bold font indicates a statistically significant difference between 2021 and 2014.

