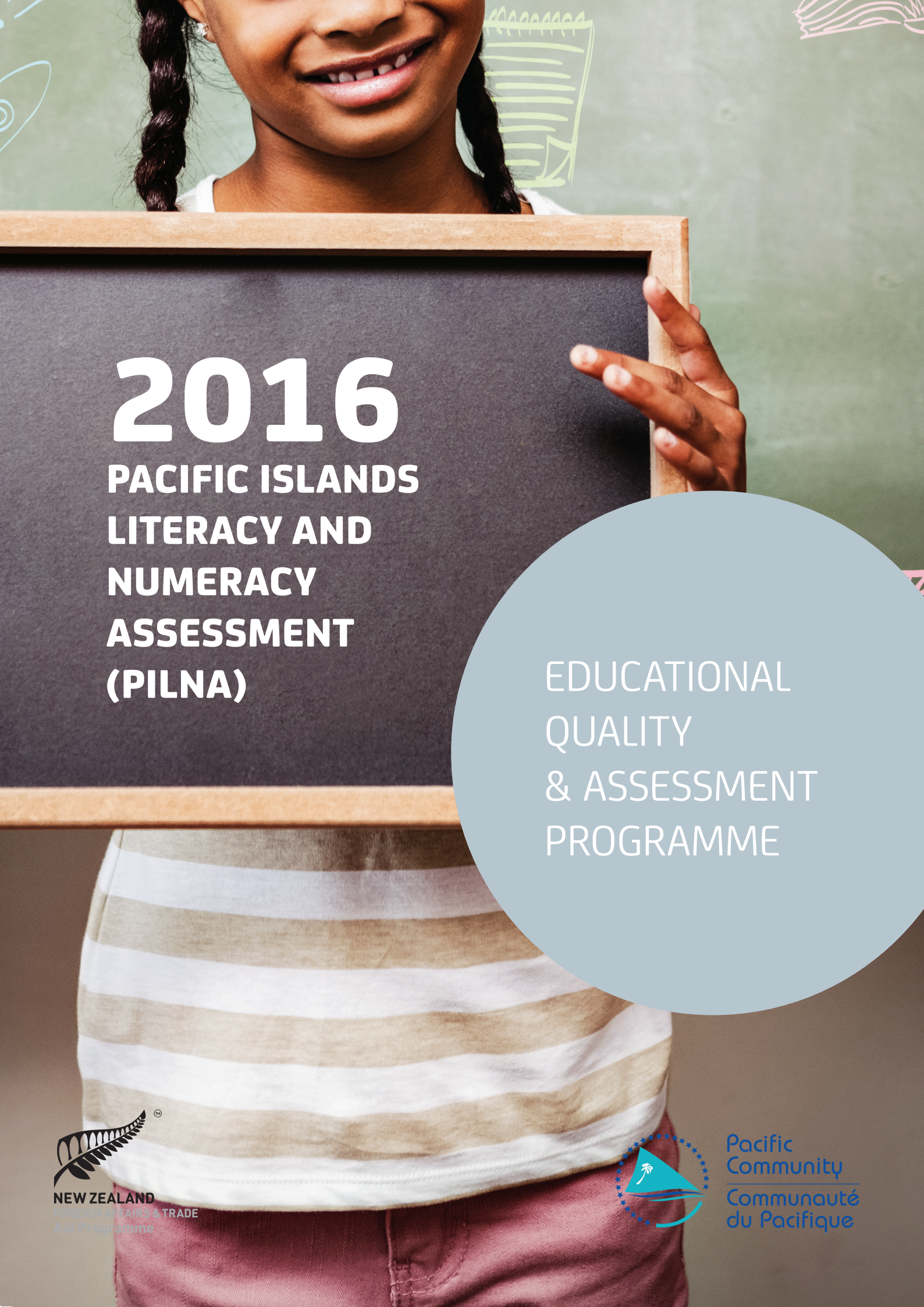


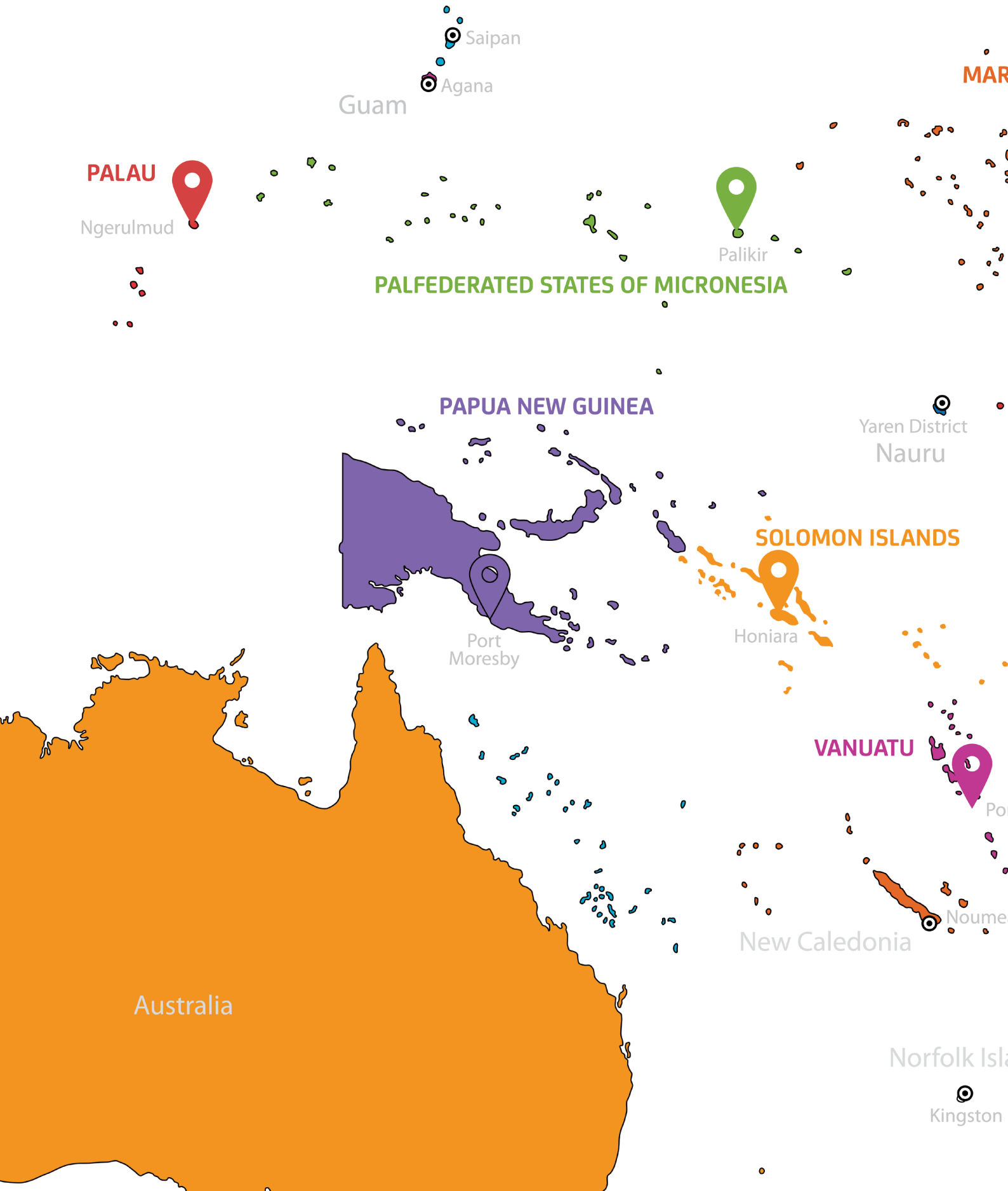
2016
PACIFIC ISLANDS LITERACY AND NUMERACY ASSESSMENT (PILNA)

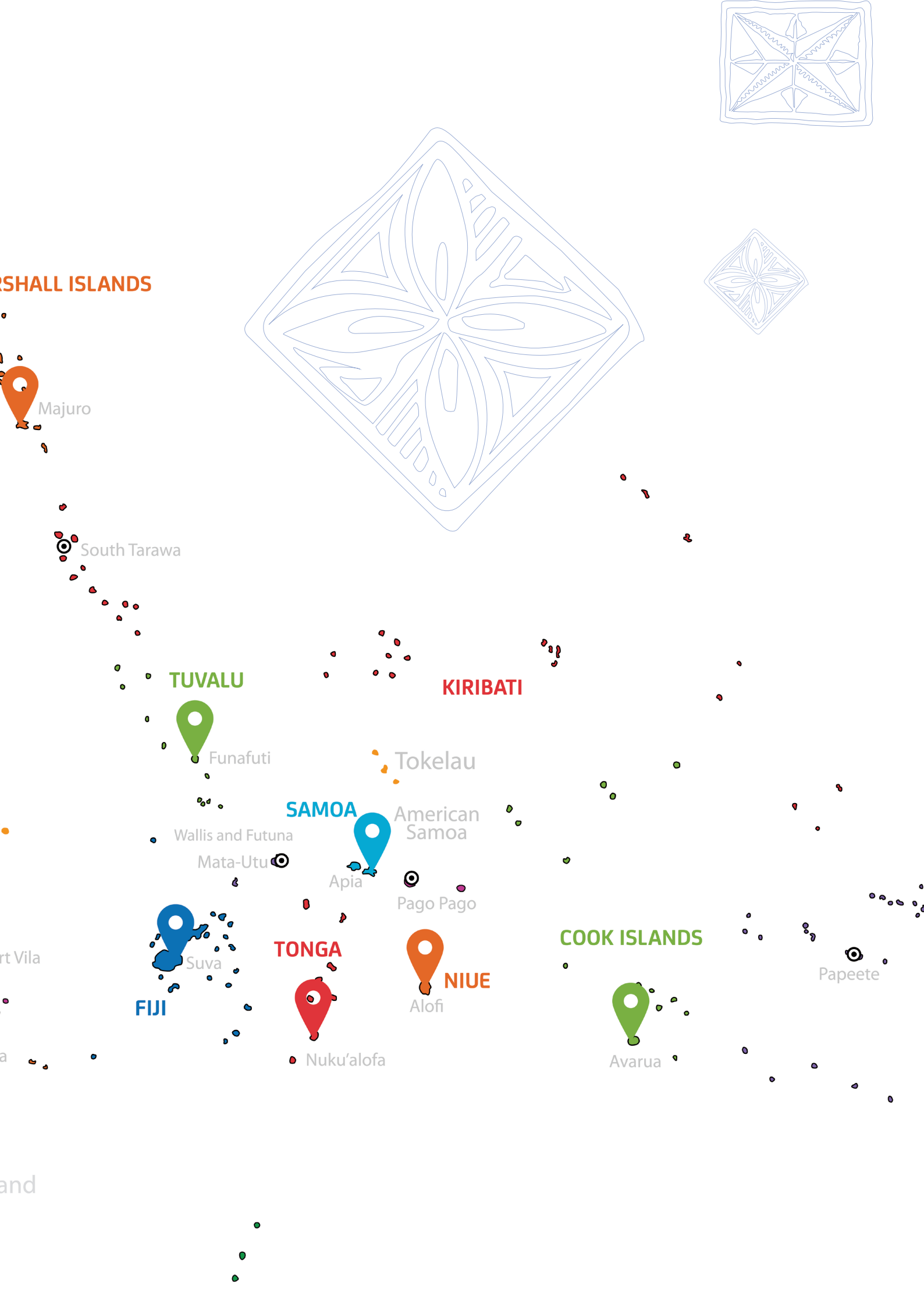
EDUCATIONAL QUALITY & ASSESSMENT PROGRAMME

NEW ZEALAND
FOREIGN AFFAIRS & TRADE
Aid Programme

Pacific Community
Communauté du Pacifique

PILNA PRESENCE




RSHALL ISLANDS
Majuro
South Tarawa
TUVALU
Funafuti
KIRIBATI
Tokelau
SAMOA
American Samoa
Wallis and Futuna
Mata-Utu
Apia
Pago Pago
Suva
FIJI
TONGA
Nuku'alofa
NIUE
Alofi
COOK ISLANDS
Avarua
Papeete
rt Vila
and

CONTENTS

ABBREVIATIONS **5**
FOREWORD **6**
EXECUTIVE SUMMARY **6**

1 Chapter 1: Introduction to PILNA 9

1.1 The Pacific context 9
1.2 Purposes of PILNA 10
1.3 PILNA 2015 10
1.4 Outline of the regional report on PILNA 2015 10

2 Chapter 2: Methodological framework 14

2.1 Data collection instruments 14
2.2 Sampling 14
2.3 Translation 17
2.4 Administration 17
2.5 Scoring and coding 18
2.5.1 The 2015 scoring/coding scheme 18
2.5.2 Data capture 19
2.5.3 Data analysis 19
2.5.4 Regional uniform metric 21
2.6 Limitations/Challenges 26

3 Chapter 3: Performance of Year 4 and Year 6 students in numeracy 27

3.1 General information on student numbers 27
3.2 Overall performance in numeracy 27
3.3 Comparing Year 4 and Year 6 27
3.4 Comparing 2012 and 2015 30
3.5 Year 4 and Year 6 domain and strands performance in numeracy 31
3.6 Year 4 and Year 6 numeracy performance by gender 32
3.7 Year 4 and Year 6 numeracy performance by school authority 35
3.8 Conclusions 37

4 Chapter 4: Performance of Year 4 and Year 6 students in literacy 39

4.1 General information on student numbers 39
4.2 Overall performance in literacy 39
4.3 Comparing Year 4 and Year 6 39

4.4 Comparing 2012 and 2015 42
4.5 Year 4 and Year 6 domain and strands performance in literacy 43
4.6 Year 4 and Year 6 literacy performance by gender 43
4.7 Year 4 and Year 6 literacy performance by school authority 47
4.8 Conclusions 50

5 Chapter 5: Coding the cognitive instrument in PILNA 2015 51

5.1 Advantages of the coding scheme 51
5.2 The coding process 51
5.3 Examples of coding in numeracy and literacy 52
5.3.1 Numeracy 53
5.3.2 Literacy 56
5.4 Conclusions 58

6 Chapter 6: Pilot questionnaire development for collection of student and school data 60

6.1 Research questions and rationale 60
6.2 Development of pilot questionnaires 61
6.3 The PILNA pilot questionnaires 62
6.4 Pilot sample and administration 62
6.5 Analysis of pilot questionnaires 62
6.6 Limitations in presenting results 63
6.7 Student questionnaires 63
6.8 Head teacher questionnaires 64
6.9 Teacher questionnaires 67
6.10 Conclusion: Use of the pilot questionnaire results for PILNA 70

7 Chapter 7: Conclusions, summary and recommendations 72

7.1 Methodological framework 72
7.2 Review of the key findings 73
7.3 Recommendations 75

References 78
Appendix 79

LIST OF TABLES

Table 2.1: Cases for numeracy by year level 14
Table 2.2: Cases for literacy by year level 14
Table 2.3: Countries with the target language for translation and documents that were translated. 15
Table 2.4: Numeracy proficiency level descriptors 20
Table 2.5: Literacy proficiency level descriptors 22
Table 3.1: Student numbers, PILNA 2015 numeracy assessment 25
Table 3.2: Distribution of students by year and proficiency levels, PILNA 2015 25
Table 3.3: Distribution of students by year and proficiency level in PILNA 2012 and PILNA 2015 28
Table 3.4: Performance in numeracy by strand and overall trend, PILNA 2012 and 2015 30
Table 3.5: Distribution of students by gender, year level and proficiency level, PILNA 2015 30
Table 3.6: Performance by gender in overall numeracy and by strand, PILNA 2015 32
Table 3.7: Distribution of students by school authority, year level and proficiency level, PILNA 2015 33
Table 3.8: Performance in overall numeracy and strands by year level and school authority, PILNA 2015 34
Table 4.1: Student numbers, literacy 2015 PILNA 36
Table 4.2: Distribution of students by year and proficiency level, PILNA 2015 37
Table 4.3: Distribution of students by year and proficiency levels between 2012 and 2015 40
Table 4.4: Performance in literacy by strands and overall trends 41
Table 4.5: Distribution of students by gender, year level and proficiency level, PILNA 2015 42
Table 4.6: Performance by gender in overall literacy and by strands, PILNA 2015 43
Table 4.7: Distribution of students by school authority, year level and proficiency level, PILNA 2015 44
Table 4.8: Performance by school authority in overall literacy and by strands, PILNA 2015 45
Table 5.1.1: Examples of the analysis for numeracy items. 50
Table 5.2: Summary of item analysis for selected literacy items 52

LIST OF FIGURES

Figure 2.1: Item-person map for numeracy 19
Figure 2.2: Item-person map for literacy 20
Figure 3.1: Regional numeracy proficiency levels, Year 4 and Year 6, PILNA 2015 27
Figure 3.2: Regional proficiency levels by year, PILNA 2015 28
Figure 3.3: Distribution of numeracy scores for Year 4 and Year 6, PILNA 2015 29
Figure 3.4: Regional numeracy proficiency levels, Year 4, PILNA 2012 and 2015 30
Figure 3.5: Regional numeracy proficiency levels, Year 6, for 2012 and 2015 30
Figure 3.6: Regional numeracy proficiency levels, Year 4 girls and boys, PILNA 2015 32
Figure 3.7: Regional numeracy proficiency levels, Year 6 for girls and boys, PILNA 2015 32
Figure 3.8: Distribution of numeracy scores by gender for Year 4 and Year 6, PILNA 2015 33
Figure 3.9: Regional numeracy proficiency levels, Year 4, by school authority type, PILNA 2015 34
Figure 3.10: Regional numeracy proficiency levels, Year 6, by school authority type, PILNA 2015 35
Figure 3.11: Distribution of numeracy scores by school authority for Year 4 and Year 6, PILNA 2015 36
Figure 4.1: Regional literacy proficiency levels, Year 4 and Year 6, PILNA 2015 39
Figure 4.2: Regional proficiency levels by year, PILNA 2015 40
Figure 4.3: Distribution of literacy scores for Year 4 and Year 6, PILNA 2015 40
Figure 4.4: Regional literacy proficiency levels, Year 4, for PILNA 2012 and 2015 41
Figure 4.5: Regional literacy proficiency levels, Year 6, for PILNA 2012 and 2015 42
Figure 4.6: Regional literacy proficiency levels, Year 4 girls and boys, PILNA 2015 43
Figure 4.7: Regional literacy proficiency levels, Year 6 girls and boys, PILNA 2015 43
Figure 4.8: Distribution of literacy scores by gender for Year 4 and Year 6, PILNA 2015 44
Figure 4.9: Regional literacy proficiency levels, Year 4, by school authority type, PILNA 2015: 45
Figure 4.10: Regional literacy proficiency levels, Year 6, by school authority type, PILNA 2015 46
Figure 4.11: Distribution of literacy scores by school authority for Year 4 and Year 6, PILNA 2015 47
Figure 5.1: Sample coding scheme 50

ABBREVIATIONS

ACER	Australian Council for Educational Research
EQAP	Educational Quality and Assessment Programme
FEdMM	Forum Education Ministers' Meeting
PILNA	Pacific Islands Literacy and Numeracy Assessment
PVs	Plausible Values
SDGs	(United Nations) Sustainable Development Goals
SIS	Small Island States
SPBEA	South Pacific Board for Educational Assessment (now EQAP)
UNESCO	United Nations Education, Scientific and Cultural Organization
UNICEF	United Nations Children's Education Fund

FOREWORD

EXECUTIVE SUMMARY

Improving educational achievement in literacy and numeracy in Pacific Island countries has been identified as a shared goal by a range of stakeholders. The Pacific Islands Literacy and Numeracy Assessment (PILNA) is a measurement of regional standards based on a common scale; it is a regional collaborative model that is highly consensual among the participating countries, providing shared intellectual capital and value for money. PILNA provides data on literacy and numeracy skills of students who have completed four and six years of formal primary education. In 2015, 13 Pacific Island countries participated in the second administration of PILNA.[1]

The Pacific is one of the largest and most diverse regions in the world, yet many countries have identified common education challenges, particularly in literacy and numeracy. Each country recognises the right of the child to have access to good quality education – of which literacy and numeracy are an inherent part – regardless of gender, ethnicity, family background or socio-economic status.

The first administration of PILNA took place in 2012 and was intended to provide a one-time snapshot of literacy and numeracy achievement in the Pacific region. Based on the insights that emerged from the findings of PILNA 2012, the Forum Education Ministers Meeting (FEdMM) requested a 2015 administration of PILNA, in addition to exploring the possibility of developing a long-term regional assessment, structured to provide valid and reliable results over time.

This commitment of FEdMM is directly linked to the United Nations Sustainable Development Goals (SDGs). By providing a measure of the literacy and numeracy skills of students who have completed four and six years of basic education, PILNA addresses targets identified in SDG 4 by providing evidence of education quality for governments, schools, communities and students in the region. Such evidence provides valuable information for stakeholders to develop interventions and policies, as well as to encourage political support and community awareness in order to improve the learning outcomes of young people in the Pacific.

Key findings

The results in this report, and any differences reported between student groups, are raw results and differences, unadjusted for any relevant background factors that might explain the reported observations. Further analysis could usefully be conducted to identify relevant contributing factors, and until that is done, care should be taken in the way the results are interpreted.

1 Thirteen countries participated in PILNA 2015: Cook Islands, Federated States of Micronesia, Kiribati, Marshall Islands, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu and Vanuatu.

Notable improvement in numeracy achievement across the region as students progress from Year 4 to Year 6 between 2012 and 2015.

Student performance in numeracy achievement improved both in student progression between year levels and in the proportion of students performing at higher proficiency levels. Students also demonstrated improved performance in all strands of the numeracy domain – numbers, operations, and measurement and data – at both year levels between 2012 and 2015.

Girls demonstrated higher levels of performance in numeracy across the region in 2015.

There were more girls than boys in the higher proficiency levels in both Year 4 and Year 6 and more boys than girls in the lower proficiency levels. Girls also outperformed boys in all strands of the numeracy domain. These findings aside, however, the difference in mean performance between girls and boys was small.

The performance in numeracy of students in government and non-government schools varied between year levels in 2015.

Based on the average regional achievement in numeracy, in Year 4 students in non-government schools slightly outperformed students in government schools. This finding is reversed in Year 6, where students in government schools slightly outperformed those in non-government schools.

Some improvement in literacy in the distribution of students between proficiency levels from 2012 to 2015.

Improvement in literacy across the region was inconsistent. There was a small improvement in performance as students progressed from Year 4 to Year 6, with more students performing at higher proficiency levels and fewer students performing at lower proficiency levels. The average mean performance in literacy of Year 4 students improved between 2012 and 2015, but there was no change for Year 6 students between the two cycles. Students in both year levels performed best in the strand of reading, followed by language features and writing. Girls demonstrated higher levels of performance in literacy across the region in 2015.

There were more girls than boys in the higher proficiency levels in both Year 4 and Year 6 and more boys than girls in the lower proficiency levels. Girls also outperformed boys in all strands – reading, language features and writing – of the literacy domain. On average, Year 6 boys were performing at a similar level to Year 4 girls across the region.
In literacy, students in non-government schools outperformed students in government schools in both year levels in 2015.

The comparative performance in literacy of students in government and non-government schools showed a uniform increase from Year 4 to Year 6, but any difference between school types was relatively small. Across the literacy domain in both types of school authority, students performed best in the reading strand, followed by the language features and writing strands.

In conclusion, PILNA 2015 has continued to build an evidence base on student learning outcomes in literacy and numeracy in the Pacific region. PILNA is an ongoing programme that can offer insights for education policy and practice, and can support the monitoring of trends in students' acquisition of knowledge and skills in literacy and numeracy. The development of PILNA 2015, the methodology used and the findings are reported in this regional report, as well as in individual country reports and a report covering the small island states. Each report draws conclusions from the findings and makes recommendations based on the evidence. In this way, PILNA addresses its ultimate aim which is to support the improvement of numeracy and literacy skills of children in the Pacific region.

Recommendations

- Educational stakeholders are advised to review PILNA evidence and trends between 2012 and 2015 both regionally and nationally, and consider intervention strategies for students performing at the lower end of the proficiency scale, particularly in literacy.
- To make certain that results reach the classroom level for targeted intervention, education authorities are advised to expand their dissemination approaches when reporting the results of the study, making certain that results reach the classroom for targeted intervention.
- Education stakeholders are strongly encouraged to identify intervention strategies that improve the achievement of boys, especially in literacy.
- Education stakeholders and EQAP are strongly encouraged to adopt the implementation of context questionnaires as part of a long-term assessment programme.
- Education stakeholders and EQAP are strongly encouraged to adopt the implementation of a full coding scheme.
- Regional and national education leaders and FEdMM are strongly encouraged to adopt the use of a regional uniform metric as a way to track progress and trends in student learning outcomes.
- Regional education stakeholders are strongly encouraged to support an ongoing PILNA that has the power to provide policymakers with more robust evidence and richer data from which to develop policies and intervention strategies to improve student learning outcomes.
- Education stakeholders are advised to investigate ways in which the robust and valid data provided by PILNA can support the improvement of student learning outcomes.

CHAPTER 1: INTRODUCTION TO PILNA

1.1 THE PACIFIC CONTEXT

The Pacific Island region is one of the largest and most diverse regions of the world and the Pacific Ocean is the world's largest body of water. Over 30 Pacific Island countries and territories are scattered throughout the region, which is home to 9.7 million inhabitants, 90 per cent of whom live in Fiji, Papua New Guinea and Solomon Islands (UNESCO 2015), while six countries have populations of less than 20,000 people. The region is characterised by rapidly changing economic structures, high migration rates and high youth unemployment in many areas. As a result of climate change, Pacific Island environments are increasingly fragile and prone to natural disasters (UNESCO 2015). While there are significant differences in geography, population and resources in the region, many countries have identified common education challenges, particularly in literacy and numeracy.

Improving educational achievement in literacy and numeracy in Pacific Island countries has been identified as a shared goal by a range of stakeholders. They recognise the right of the child to have access to good quality education – of which literacy and numeracy are an inherent part – regardless of gender, ethnicity, family background or socio-economic status. Pacific leaders are cognisant of international studies that have highlighted the relationship between literacy and numeracy skills and full participation in society (OECD 2014; Altinok 2012; Duncan, et. al. 2007; Lewin 2007). More critically, Pacific leaders are looking at ways to reverse the global trend of many young people, especially the disadvantaged, leaving school without the skills to engage in everyday society and secure employment (UNESCO 2012). Pacific Island stakeholders understand that literacy and numeracy are foundation skills necessary to engage with all aspects of everyday life.

In 2006, 15 Pacific countries collaborated to develop Pacific-wide benchmark standards for literacy and numeracy at years 2, 4, 6 and 8.[2] The benchmarks were developed from curriculum skill components and learning outcomes determined to be common across the national curricula of Pacific Island countries.

The consultations resulted in the document Setting regional benchmarks for literacy, numeracy and life-skills to monitor the quality of basic education in the Pacific region, which was endorsed at the Forum Education Ministers Meeting in 2007.This document is the basis for monitoring the quality of education by assessing literacy and numeracy levels in the Pacific region.

The regional benchmarks document defines literacy and numeracy, indicating what a literate or numerate person is in the Pacific context. The definition of literacy (page 3) is:

"Knowledge and skills necessary to empower a person to communicate through reading and writing, in particular language or languages, with respect to their society and individual needs."

The definition of numeracy (page 4) is:

"Knowledge and skills necessary to empower a person to be able to use numbers in mathematical processes, as well as the language of mathematics, for a variety of purposes, with respect to everyday life."

1.2 PURPOSES OF PILNA

A regional assessment programme provides an important tool for collecting comparative and benchmarking[3] data. The Pacific Islands Literacy and Numeracy Assessment (PILNA) is a measurement of regional standards based on a common scale; it is a regional collaborative model that is highly consensual among the participating countries, providing shared intellectual capital and value for money. PILNA provides data on literacy and numeracy skills of students who have completed four and six years of formal primary education.

With the prioritisation of student learning outcomes, countries elected to focus on six key areas for the use of data and reports resulting from PILNA. These six areas are: using data for system- and school-level interventions; using evidence to inform policy development; using data for political support of assessment for learning; using the findings to promote community awareness about learning outcomes; monitoring of results; and using PILNA data to validate national assessment processes.

2. Support for the benchmark development was provided by UNESCO and EQAP (formerly the South Pacific Board for Educational Assessment [SPBEA]).
3. Benchmarking is a process that enables the measurement of students' academic growth based on an agreed set of learning outcomes indicators. The benchmarking process documents what students have demonstrated to reach the minimum level of knowledge and skills expected at a particular school year level. PILNA is linked to the Pacific regional benchmarks document of 2007.

1.3 PILNA 2015

In 2012, the Pacific Islands Literacy and Numeracy Assessment (PILNA) was administered across 14 Pacific Island countries as a one-time snapshot of literacy and numeracy levels in the region.[4] Once data analysis and reporting were finalised, the results of PILNA 2012 were presented to the 2014 Forum Education Ministers Meeting (FEdMM). The findings provided an insight into student achievement in literacy and numeracy across the region, and the results were such that FEdMM requested a 2015 administration of PILNA.[5]

The ministers also recommended exploring the possibility of developing a long-term regional assessment, structured to provide valid and reliable results over time, in order to support existing efforts to improve educational outcomes.

This commitment of FEdMM is directly linked to the United Nations Sustainable Development Goals (SDGs), which outline a global commitment to a 15-year agenda to tackle poverty through initiatives that encompass the environmental, social and economic dimensions of sustainable development (UNDP 2015). SDG 4 specifically focuses on quality education and provides a framework for PILNA.
By providing a measure of the literacy and numeracy skills of students who have completed four and six years of basic education, PILNA addresses targets identified in SDG 4 by providing evidence of education quality for governments, schools, communities and students in the region. Such evidence provides valuable information for stakeholders to develop interventions and policies, as well as to encourage political support and community awareness in order to improve the learning outcomes of young people in the Pacific.

1.4 OUTLINE OF THE REGIONAL REPORT ON PILNA 2015

This chapter introduces PILNA, including a description and purpose of the assessment. It needs to clarified that the results in this report, and any differences reported between student groups, are raw results and differences, unadjusted for relevant background factors that might explain the reported observations. This applies to all the comparisons in the regional, Small Islands States and country reports. Further analysis could usefully be conducted to identify relevant contributing factors and, until that is done, care should be taken in the way the results are interpreted.

Chapter 2 provides an overview of the methodological framework, data analysis and the development of a common scale and proficiency benchmarks. All results are presented at the regional level and on a regional scale.

Chapter 3 addresses the performance of Year 4 and Year 6 students in numeracy. It begins by discussing students' overall numeracy performance in the region, and then provides a picture of numeracy achievement in the domain and subscales, or strands, of the domain. For numeracy, these strands are numbers, operations, and measurement and data. The chapter then explores performance by gender and by school authority (grouped as government and non-government schools).

Chapter 4 addresses the performance of Year 4 and Year 6 students in literacy. It begins by discussing students' overall literacy performance in the region, and then provides a picture of literacy achievement in the domain and subscales, or strands, of the domain.
For literacy these strands are reading, language features and writing. The chapter then explores performance by gender and by school authority (grouped as government and non-government schools).

The results for Chapter 3 and Chapter 4 are presented in the following formats:

- proficiency level tables and histograms;
- tables of descriptive statistics for the domain and subscale scores; and
- box plots. See inset below for an explanation on how to read and interpret box plots.

Chapter 5 explores the development of a coding scheme for PILNA. A coding scheme enables a more in-depth item analysis. Examples of how coding was partially implemented for PILNA 2015 are also included in this chapter. The information from this analysis presents opportunities for teaching interventions, as classroom teachers can use it to address misconceptions by students on specific topics.

Chapter 6 discusses the development and administration of pilot questionnaires for students, teachers and head teachers. As outlined in Chapter 2, a questionnaire is a key part of any long-term assessment programme. This chapter also addresses some indicative results of the pilot and steps for refinement of the questionnaires.

Chapter 7 summarises the major conclusions of PILNA 2015. It also provides recommendations on potential next steps for future cycles.

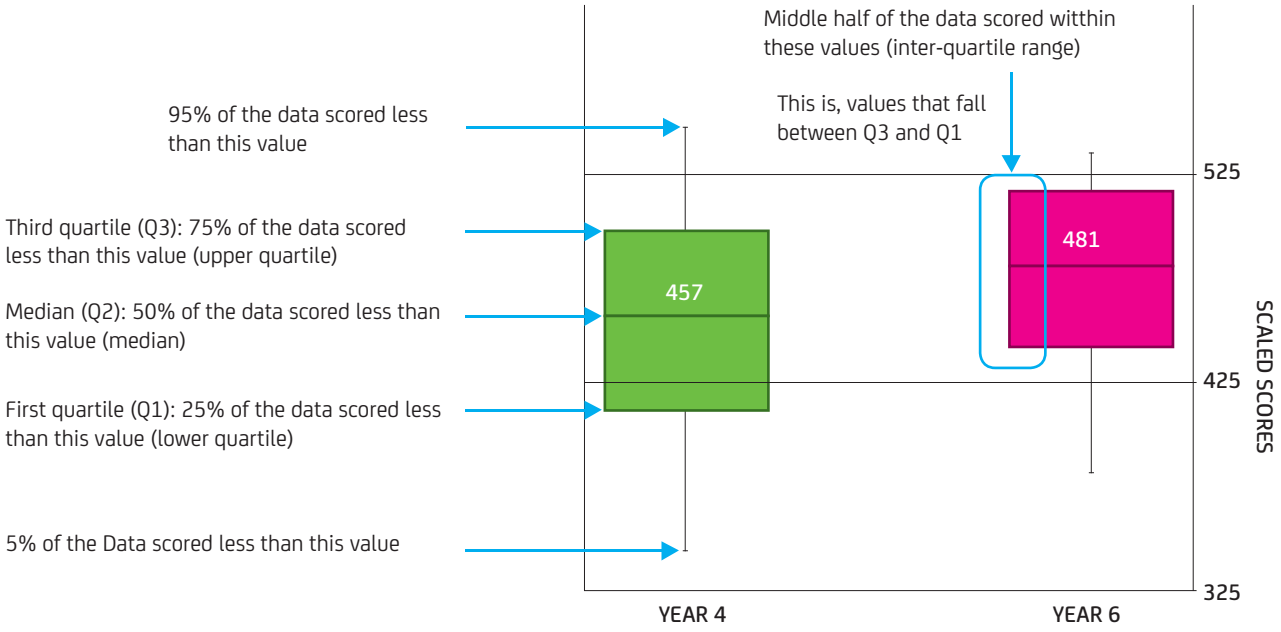
4. The 14 countries that took part in PILNA 2012 are: Cook Islands, Federated States of Micronesia, Fiji, Kiribati, Marshall Islands, Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tuvalu and Vanuatu.
5. Thirteen countries participated in PILNA 2015: Cook Islands, Federated States of Micronesia, Kiribati, Marshall Islands, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu and Vanuatu.

BOX PLOT

What is a box plot?

A box plot summarises a large amount of data graphically, displaying the distribution of data along a scale. Box plots have the advantage of enabling users to compare a number of datasets or subgroups within a dataset at one time on a common scale, making differences between them readily apparent.

Figure X shows two box plots based on PILNA data for the literacy domain for Years 4 and 6. Each of the box plots has four parts – two adjoining boxes in the middle, and a whisker[6] extending from each side of the middle boxes.



6 A box plot is sometimes referred to as a box and whisker plot

We could imagine that all data points (for example, the scores for all students in Year 4) are lined up in order from smallest to largest, then divided into four equal groups. We refer to the boundaries between those four equal parts of the distribution as the quartiles, since they define the location of the four quarters of the distribution. The boundaries are referred to as:

Q1 - the boundary between the lowest quarter and the second quarter, it marks the score that is one-quarter or 25% of the way along the ordered scores, and so is sometimes referred to as the 25th percentile;

Q2 - the boundary between the two middle quarters – this middle point of the distribution also has a special name: the 'median', and is sometimes referred to as the 50th percentile;

Q3 - the boundary between the third quarter and the highest scoring quarter, also referred to as the 75th percentile.

Box plots display the two middle quarters in two boxes, with their boundary (the median) being labelled as a particular score point. Above the and below those boxes are two 'whiskers', which are single lines extending upwards from the third quartile, and downwards from the first quartile. This particular version of the box plot uses whiskers that extend upwards from the 75th percentile (Q3) to the 95th percentile, and downwards from the 25th percentile (Q1) to the 5th percentile. This means that the highest and lowest 5% of scores are not included in the representation. This can be useful, since outliers can distort data representations of this kind. The box plot, therefore, captures the middle 90% of the distribution, omitting only the extreme values at each end.

In the example above the whiskers tell us a very low Literacy domain score for our sample of Grade 4 and 6 students (only 5% of scores are lower) and a very high Literacy domain score for Year 4 and 6 students (only 5% of scores are higher). In Figure X ninety percent of the scores for Country X in literacy range between 335 points and 540 points for Year 4 and 380 points and 530 points for Year 6. Only the very few extreme scores lie outside these ranges.

Q2 and Q3 define the edges of the box component of the box plot. The line through the middle of the box is the median score for the entire dataset. Half of the scores lie above this point, and half lie below. In the example above, the median score for Year 4 is 457 points and the median score for Year 6 is 481 points. The top of the box is at Q3, and the bottom of the box is at Q1. These scores mark the top and bottom scores for the middle half of the dataset (the two middle quarters). In the example, the third quartile (Q3) for Year 4 is 490 points and Q1 for Year 6 is 515 points; while the first quartile (Q1) for Year 4 is 420 points, and for Year 6 is 445 points. Remember, each pair of adjacent quartiles surrounds 25% of the dataset. If one side of the box is longer than the other, it does not mean that side contains more data. Rather, it means the same number of scores are spread out over a greater part of the score scale.

Why is a box plot useful?

A box plot is useful as it tells the reader the spread and mid-point of a dataset. Using the box plot for Year 4 in Figure X as an example, the box plots tell us that Country X has Year 4 students who achieved domain scores of 540 points (this is the highest proficiency level for literacy – see Section 2), and that only the highest-performing 5% of students scored higher. However, the median for Year 4 is 457 points which is classified as Level 5 proficiency. The box plot tells us that, on average, Year 4 students in Country X are performing well in literacy as they are performing at the expected level for Year 6. It also tells us that students below the top quarter of the population have scores that are clustered across a smaller point score range. However, students below the lower quartile have a wider range of scores (as depicted by the longer whisker below Q1).

In addition, putting two box plots side by side also allows for the comparison of the distribution between two groups (e.g. between Year 4 and Year 6 in Figure X). Figure X shows that the range of scores for Year 4 is much wider than the range for Year 6, such that some Year 4 students achieved higher scores than Year 6 students. Figure X also shows that, while the spread of scores for Year 6 is narrower compared to that of Year 4, some students still lag behind the majority of their peers.

CHAPTER 2: METHODOLOGICAL FRAMEWORK

2.1 DATA COLLECTION INSTRUMENTS

Consistent with a high quality learning assessment programme, two data collection components were developed for PILNA 2012: a cognitive component and a questionnaire component. Each component is discussed from a methodological perspective.

The cognitive component covers the subject matter that is being assessed. Depending on the nature of the sampling used, this component gives information about how students perform on the test questions. Importantly, this component gives comparative information about learning outcomes for different students, classes, schools, regions and countries.

Quantitative research instruments were developed for PILNA 2012. They were designed to provide reliable and valid data on the achievement levels in the literacy and numeracy skills of students who had completed Years 4 and 6. After PILNA 2012, reviews of the data and the instruments were undertaken, which led to the development of the 2015 instruments.

The questionnaire component of an assessment includes the collection of background and contextual data at different levels. The information gathered can be a powerful tool in providing explanations for the outcomes of an assessment's cognitive component. This information enables a more in-depth understanding of the observed test outcomes (student learning outcomes), and the implications of these outcomes for designing improvement strategies.

EQAP designed three pilot questionnaires which were administered to students, teachers and head teachers. The questionnaires were framed around the following research question:

What factors influence students' achievement in PILNA?
To address this research question, the following sub-questions were included to guide the questionnaire development for students, teachers and head teachers.

a. How do socio-economic factors influence student performance in PILNA?
b. How do student study habits influence achievement in PILNA?
c. What influence does teacher quality have on student achievement in PILNA?
d. How does school management and leadership influence student achievement in PILNA?

The questionnaires were piloted in 2015 with the aim of providing empirical evidence about the relevance, reliability and potential usefulness of the questions selected for inclusion.

The overall methodology of PILNA provides a comparative analysis of data with the Pacific regional benchmarks, student performance on PILNA 2012, and student performance of countries in the region as a whole. It is important to note that country-to-country comparison is NOT a component of the project, as explicitly directed by FEdMM in 2014.
Thirteen countries participated in PILNA 2015: Cook Islands, Federated States of Micronesia, Kiribati, Marshall Islands, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu and Vanuatu.

2.2 SAMPLING

Given the extreme variations that apply across the countries participating in PILNA, the sampling design is a complex process. The design uses a census approach for the relatively smaller countries of Cook Islands, Niue, Palau, Tokelau and Tuvalu, and a sampling approach for Federated States of Micronesia, Papua New Guinea, Republic of the Marshall Islands, Samoa, Solomon Islands, Tonga and Vanuatu. Although it is a comparatively larger country population-wise, Kiribati requested that a census approach be applied to its PILNA administration.[7]

In the sampling frame, schools are categorised by:

- country, education district/province/region, location address, locality (urban, non-urban, remote, very remote)
- school authority (government, private, religious authority)
- regional authority (government, non-government)school
- size, Year 4 size and Year 6 size.

These variables are available for analysis purposes where the numbers permit such analysis. The region/district/province variable is used as the basis of a statistically robust, stratified random sampling approach[8] for the sampled countries.

The target population is defined as Year 4 and Year 6 students (Year 5 and Year 7 for countries in the northern Pacific area). It is important to note that all students, regardless of year as defined in each country, have completed four or six years of equivalent schooling. It is clear that the definition for the targeted population for PILNA 2015 is based entirely on year level description and not on the age of students. Excluded parts of the population include Year 4 and Year 6 students attending special schools, students in very remote and isolated schools, students in very small schools, and students in schools offering an international curriculum.

For the 2015 PILNA, the basic sampling approach can be summarised as approximately 2,000 students selected from a two-stage stratified sampling frame:

- approximately 93 schools selected with probability proportional to size; and
- one intact class of students sampled from each school, yielding on average 25 students per school (before non-response).

The sampling plan assumes an intra-class correlation coefficient (ICC) of 0.2. Under this assumption, the plan states that a sample of 2,000 students selected under this design with students clustered in class groups of about 25 would be roughly equivalent with respect to the precision of estimates to a simple random sample of 400 students. This conforms to standard sampling practice.

The main stratification variable is the administrative divisions of the country. After all exclusions were removed, the national sampling frame was grouped into strata. Each stratum[9] represented a principal administrative division found at the sub-national level within the sampled countries.

Table 2.1 lists the number of students who took the PILNA numeracy test in 2015. Students are listed by country and by year level.

7 The sampling process is documented in detail in the 2015 PILNA Technical Report.
8 A stratified random sample is a population sample that requires the population to be divided into smaller groups, called 'strata'. Random samples can be taken from each stratum, or group.
9 Strata: province, region, district or island

Table 2.1: Students taking the numeracy test by year level

COUNTRY	YEAR		TOTAL
	YEAR 4	YEAR 6	
Cook Islands	253	214	467
FSM (Micronesia)	1360	1325	2685
Kiribati	2152	1828	3980
Niue	27	27	54
Palau	233	218	451
Papua New Guinea	0	3051	3051
Marshall Islands	700	612	1312
Samoa	1818	1853	3671
Solomon Islands	1548	1672	3220
Tokelau	21	37	58
Tonga	1695	2067	3762
Tuvalu	202	176	378
Vanuatu	1280	1212	2492
Total	**11289**	**14292**	**25581**

Table 2.2 lists the number of students who took the PILNA literacy test in 2015. Students are listed by country and by year level.

Table 2.2: Students taking the literacy test by year level

COUNTRY	YEAR		TOTAL
	YEAR 4	YEAR 6	
Cook Islands	253	219	472
FSM (Micronesia)	1498	1450	2948
Kiribati	2264	1909	4173
Niue	29	30	59
Palau	247	222	469
Papua New Guinea	0	3713	3713
Marshall Islands	755	664	1419
Samoa	1860	1929	3789
Solomon Islands	1770	1903	3673
Tokelau	24	37	61
Tonga	1867	2166	4033
Tuvalu	197	195	392
Vanuatu	1430	1370	2800
Total	**12194**	**15807**	**28001**

2.3 TRANSLATION

In line with the definition of literacy in the regional benchmarks, PILNA countries were given the opportunity to consider their individual language policies and the language of instruction/testing at both Year 4 and Year 6. Nine countries opted for translated versions of the PILNA instruments. Table 2.3 shows the countries and the target language of testing for the instrument translations.

Table 2.3: Countries with the target language for translation and documents that were translated

Country		Target language	Translated instruments and instructions
1	Cook Islands	Cook Islands Maori	Y4 numeracy and test supervisor's instructions Y4 literacy and test supervisor's instructions
2	Niue	Vagahau Niue	Y4 numeracy, Y6 numeracy and test supervisor's instructions Y4 literacy and Y6 numeracy and test supervisor's instructions
3	Tonga	Tongan	Y4 numeracy, Y6 numeracy and test supervisor's instructions Y4 literacy and test supervisor's instructions
4	Vanuatu	French	Y4 numeracy, Y6 numeracy and test supervisor's instructions Y4 literacy, Y6 literacy and test supervisor's instructions
5	Kiribati	Te Kiribati	Y4 numeracy and test supervisor's instructions Y4 literacy and test supervisor's instructions
6	RMI	Marshallese	Y4 numeracy, Y6 numeracy and test supervisor's instructions Y4 literacy, Y6 literacy and test supervisor's instructions
7	Tokelau	Tokelauan	Y4 numeracy and test supervisor's instructions; Y6 numeracy
8	Tuvalu	Tuvaluan	Y4 numeracy and test supervisor's instructions; Y6 numeracy
9	Samoa	Gagana Samoan	Y4 numeracy

2.4 ADMINISTRATION

The following tools were administered for PILNA 2015 data collection:

- literacy and numeracy assessments in Year 4 and Year 6 (cognitive instruments);
- pilot student questionnaires (contextual instruments);
- pilot teacher questionnaires (contextual instruments); and
- pilot principal/head teacher questionnaires (contextual instruments).

PILNA was administered over two days in each participating country in October 2015. If administration occurred outside the month of October, the data collection window was agreed upon by the participating country and EQAP. Cognitive instruments were administered in ten different languages (discussed in the translation methodology).

Instruments were sent to each country's education office in late August 2015, and PILNA national coordinators despatched the instruments to selected schools. Packages (including cognitive instruments, pilot questionnaires and an implementation manual) were distributed to schools to allow reasonable time for the school coordinators to brief the test supervisors.

Data for PILNA 2015 were collected, scored and housed under strict security protocols. Data were validated and stored in one database for all countries.

Most of the validation processes were done in-country, using the following processes:

- panel leaders for literacy and numeracy re-checked at least ten per cent of marked scripts;
- the data entry screen had strict pre-set rules; and
- data reports were generated and printed to check entered data.
-

Validated data were then cleaned, coded and aggregated in January 2016 at EQAP. EQAP operational procedures ensured that related responses to the different instruments were linked.

2.5 SCORING AND CODING

Scoring and coding are two aspects of the treatment of students' test responses. The responses that students provide to test questions are first coded, meaning that they are assigned to pre-defined response categories that were established as part of the test development process. Scoring usually happens after that, when each response code is assigned a quantitative value (a score) and when statistical processes are applied to the scored responses to estimate the abilities of the sample.

Coding of student responses was carried out in-country. EQAP officers were in-country to train coders and to supervise the scoring and data entry. PILNA national coordinators identified a numeracy coding panel leader and a literacy coding panel leader, and appointed the members of the panels, which also included data entry officers. Panel members were selected, based on their experience with coding, as well as their content knowledge in literacy or numeracy. Data entry officers entered students' scores online or on a pre-prepared Excel spreadsheet, and questionnaire responses were scanned at EQAP in Suva, Fiji.

2.5.1 THE 2015 SCORING/CODING SCHEME

A review of the outcomes of PILNA 2012 suggested that more information about student understanding and misunderstanding could be obtained with a relatively minor adjustment to the approach to item design. As a result, preliminary steps were taken to adjust the response coding scheme for a number of the items used in the 2015 PILNA administration. The coding scheme was piloted for particular items on the literacy and numeracy instruments. The process and design of the coding scheme is discussed in detail, including examples of coded questions and responses, in Chapter 5.

2.5.2 DATA CAPTURe

PILNA uses a web-based platform for capturing data. Users are able to access the system using their username and password. It has a database that is secure and accessible only by the administrator and the application. Access to parts of the database is controlled through the use of roles that determine what type of access should be granted to the user.

The data management application has four sections:

- administration section – for configuration of the capture form;
- report section;
- online form – the online mode of the capture form; and
- template – provides the template for offline mode and also the uploading of the template file.

To ensure the integrity of the data, the system is able to validate the data entered by the user and provide the appropriate feedback on data entry in both the online portion of the application and for users entering data via the Excel templates. The system also keeps track of the user who entered the data.

2.5.3 DATA ANALYSIS

The data for both Year 4 and Year 6 were combined into a single dataset for analysis. Rasch modelling was used to scale the data for numeracy and literacy. This combined dataset was jointly calibrated using ACER ConQuest software. The joint calibration allows both years 4 and 6 to be put on a single, uniform scale. Student ability was estimated using plausible values (PVs) and PVs were generated for each domain, as well as for subscales, or strands. For both numeracy and literacy, three subscales/strands each were defined as follows:

- NUMERACY
  - Numbers
  - Operations
  - Measurement and data
- LITERACY
  - Reading
  - Language features
  - Writing

In order to link the 2012 and 2015 scales, item parameters from the 2015 calibration were used as anchors (for common items between the two PILNA cycles) to generate new student ability estimates for the 2012 test population, and these were used for the 'horizontal linking' across the two assessment cycles.

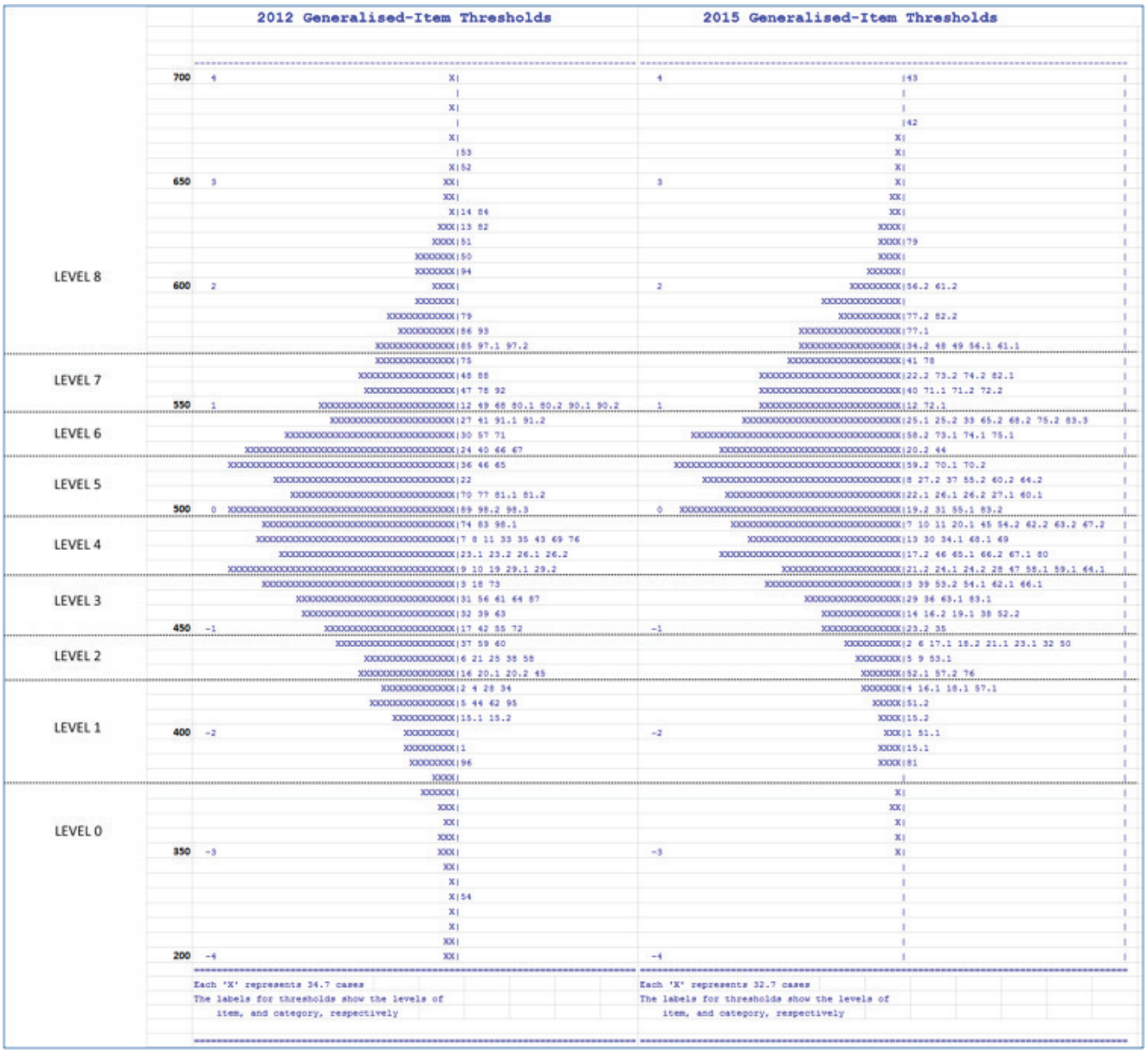


Figure 2.1: Item-person map for numeracy

Figure 2.1 represents the distributions of test item difficulties and abilities of assessed students for numeracy, with the distribution for 2012 shown on the left and for 2015 on the right, displayed on a common numerical scale. The common scale is expressed in two different units – the 'logit' unit (shown to span from -4 to +4) appears next to each of the two parts of the display, and the transformed PILNA unit (spanning from 200 to 700) is shown further to the left of the display. The defined proficiency levels are also marked as bands across the display. Data for each of the two years are spread vertically along the scale, with the symbol 'X' on the left representing the numeracy abilities of students at different points along the scale, and the number labels on the right showing where each item is located in terms of difficulty along the same scale. Students of lower ability and items of lower difficulty are at the bottom of the display, with student ability and item difficulty both increasing further up the display.

The map also shows how well the numeracy tests for the two years are targeting the test students in each cycle. This can be seen in the relationship between the distribution of students (on the left of each part of the display) and the distribution of items (on the right). The distributions are broadly aligned with each other.

The same relationships are shown for literacy in Figure 2.2.

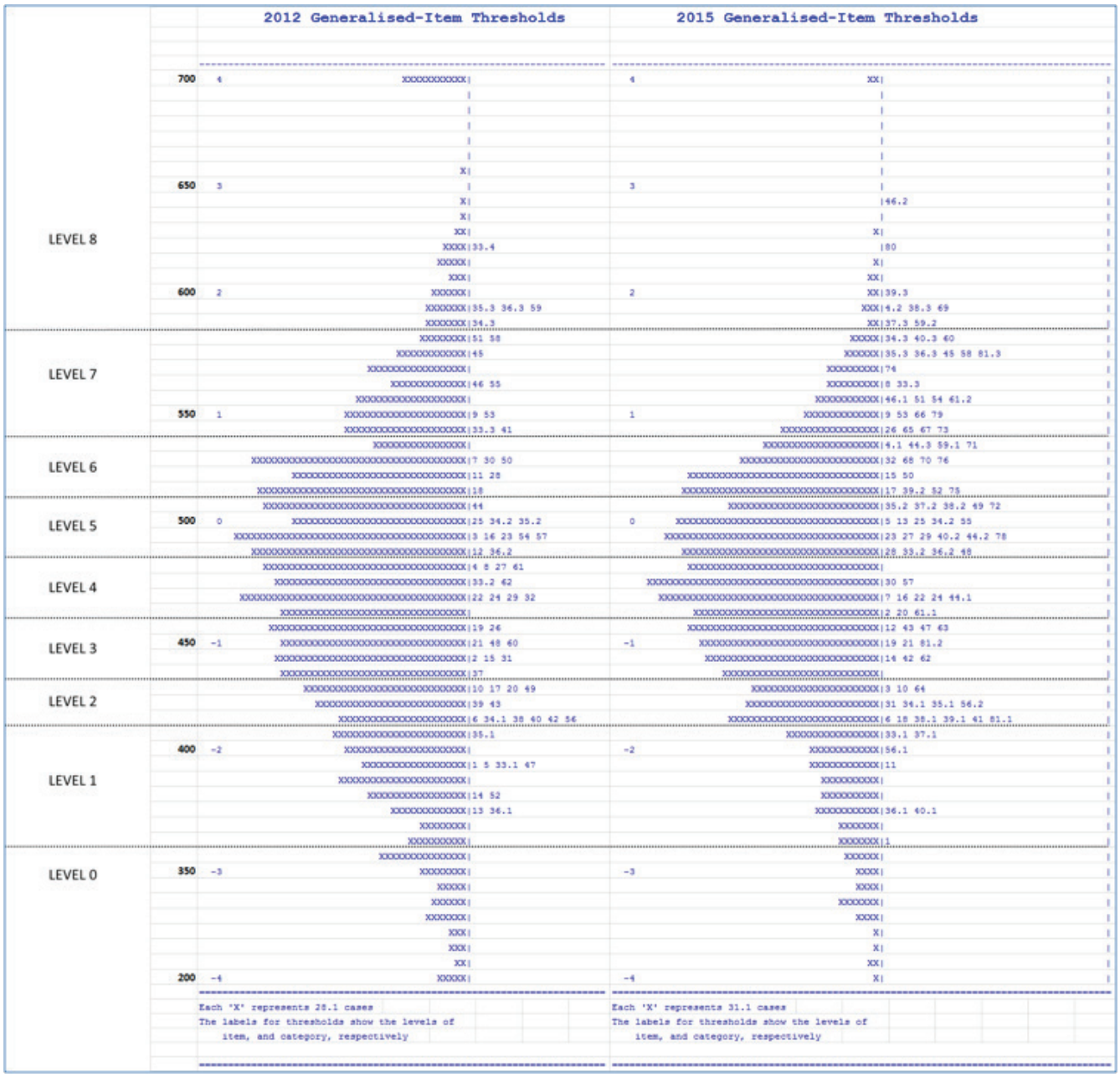


Figure 2.2: Item-person map for literacy

2.5.4 REGIONAL UNIFORM METRIC

To enable a consistent approach to reporting across all PILNA countries, student outcomes have been reported on a single uniform metric that is applied across the region. The reporting metric has been constructed to achieve two main goals: first, to provide descriptions of what students can do at various points along the metric; and second, to show results in a way that can be interpreted consistently across all participating populations. This means that results can readily be compared across different parts of each country's population (for example, across students from urban and non-urban areas, between girls and boys). National results can also be compared with the average achievement across the region.

2.5.4.1 DEVELOPMENT OF THE PROFICIENCY LEVELS

The proficiency levels were developed, and the levels described, using a process involving a panel of experts and following several steps that are summarised here.

1. A 'generalised item thresholds' table was prepared, containing all items from both 2012 and 2015 cycles. This is essentially a listing of each available score point across all items, ordered by the difficulty of obtaining each score point.
2. Descriptors for each score point were attached to the ordered list. These descriptors encapsulated the key cognitive demand or the particular skill involved in obtaining each score point.
3. These descriptors were then used to develop the summary proficiency level descriptions. The 2015 items were prioritised in deciding the level cut-offs and in developing the summary level descriptions.

The set of new proficiency scale levels was developed, based on the item-to-skill mapping and placing the items on a Guttmann structure (i.e. ordering the items based on difficulty and establishing level cut-offs based on the skill and content grouping of the items). Although this process results in levels that are not strictly of equal width in terms of item difficulty, the panel endeavoured to make the levels as uniform as possible. The summary descriptors for each proficiency level are described in Table 2.4 and Table 2.5 for numeracy and literacy respectively.

The ability estimates from the IRT analysis are originally reported in units that are called logits, with a mean[10] of 0 and standard deviation[11] of 1. To avoid the confusion that might arise from reporting negative scores, the scaled scores that will be used for public reporting have to fit in a range that does not include negative numbers. The ability estimates in logits were converted into a PILNA scaled score, with a mean of 500 and standard deviation of 50, using the following conversion formula:

PILNA Scaled Score = [(score in logits) x 50] + 500

making it wide enough for current and foreseeable future needs. The equivalence between scores in logits, the transformed PILNA scaled scores, and corresponding proficiency levels are shown visually in Tables 2.4 and 2.5.

10 The arithmetic mean, also commonly referred to as the average. The mean is the sum of all scores in a sample divided by the number of scores in that sample.

11 The standard deviation is a standardised measure of spread in a distribution (the distribution of scores in this context). It is defined as the square root of the average squared deviations from the mean.

Table 2.4: Numeracy proficiency level descriptors

LEVEL	NUMERACY DESCRIPTORS
	Students at each of the levels 1 to 8 are able to do the skills in each described level with proper guidance by the teacher, and are likely to do the skills in the preceding lower levels independently.
8 575 or greater	Round off numbers to the nearest tenth and hundredth and convert fractions to percentages and vice versa. Add and subtract fractions with denominators that are multiples. Measure and determine the perimeter of a simple shape. Show time on a clock and solve problems involving time duration. Calculate averages from data given in a bar graph.
7 550 to < 575	Represent a proportion of a whole as a fraction and round off numbers to the nearest tens and hundreds. Divide a two-digit number by a one-digit number with a remainder, and understand the order of operation by simplifying expressions involving the four operations. Solve word problems involving both addition and subtraction to the extent of calculating the total cost and change from shopping. Tell the time from an analogue clock in minutes.
6 525 to < 550	Complete an increasing number pattern that involves decimal numbers with two decimal places, and also complete a decreasing whole number pattern. Subtract up to three-digit numbers from up to four-digit numbers with regrouping, and also subtract decimal numbers with different numbers of decimal places and with regrouping. Multiply a three-digit number by a two-digit number with regrouping to the extent of solving word problems involving multiplication, calculating unit cost and calculating change from shopping. Tell the time to the quarter hour and half hour from an analogue clock. Draw a complete bar graph that will convey information from a given set of data.
5 *Expected level for Year 6 500 to < 525	Write a four-digit number involving zeros in numerals and identify place values of a two-digit number. Add and subtract fractions with the same denominators, and add two decimal numbers with different numbers of decimal places and with regrouping. Subtract a two-digit number from a three-digit number with regrouping. Multiply a three-digit with a two-digit number without regrouping, and understand and simplify brackets to determine the order of operation. Measure height.
4 475 to < 500	Read numbers on a place value number system and compare four-digit whole numbers and decimal numbers. Identify the numerator and denominator of a fraction to the extent of representing proportion of a whole as a simple percentage. Add three two-digit whole numbers with regrouping, multiply a two- or a three-digit number and a one-digit number with regrouping, and divide a two-digit by a one-digit number without remainder. Simplify expressions involving addition and subtraction and calculate total cost of three items. Identify days in a week and read with understanding data from a bar graph.
3 *Expected level for Year 4 450 to < 475	Write a four-digit number not involving zero in words and numerals. Write a three-digit number involving zero in numerals and write a four-digit number involving zero in words. Complete increasing number patterns involving decimal numbers to one decimal place in a relation and recognise money according to its value. Add two- to four-digit numbers with two- to three-digit numbers with regrouping, and add two decimal numbers with the same number of decimal places and with regrouping. Multiply a two-digit number and one-digit number with no regrouping and solve simple word problems involving subtraction. Use a ruler to draw and read a given length and tell the time to the hour only from an analogue clock.
2 425 to < 450	Write a three-digit number not involving zero in words and in numerals, and write a three-digit number involving zero in words only. Compare prices of items and calculate the total cost of two items. Subtract a two-digit number from a two- or three-digit number without regrouping and solve simple word problems involving addition. Identify hands of a clock and know the relation of days and weeks.
1 375 to < 425	Write a two-digit number not involving zero in words and in numerals, and also complete increasing number patterns in a simple relation. Add any pair of two-digit and two- or three-digit numbers without regrouping. Compare heights of data presented in a bar graph.
0 Less than 375	Students at this level are not able to do any of the skills above and/or there is insufficient evidence to indicate their ability.

Table 2.5: Literacy proficiency level descriptors

LEVELS	LITERACY DESCRIPTORS
	Students at each of the levels 1 to 8 are able to do the skills in each described level with proper guidance by the teacher, and are likely to do the skills in the preceding lower levels independently.
8 587.5 or greater	Draw valid conclusions and explain the main arguments in an authentic text on an unfamiliar subject. Demonstrate understanding and mastery in the use of language conventions. Write a story using an expanded range of well-expressed ideas that are elaborated and organised in a coherent text with full control and use of key language features.
7 537.5 to < 587.5	Derive the author's implicit intent, make inferences and interpret information from a variety of texts. Demonstrate proficiency in spelling, punctuation, grammar, syntax and vocabulary. Write a story using an expanded range of elaborated ideas that are organised in a coherent text with good control of key language features and a variety of sentence structure.
6 512.5 to < 537.5	Relate specific information to images portrayed in poems and instructional text and draw conclusions based on evidence in a story. Demonstrate general proficiency in the use of common conventions in grammar, tense and various degrees of comparison. Write a story using a range of elaborated ideas and structure in a coherent text with correct use of language features.
5 *Expected level for Year 6 487.5 to < 512.5	Read and critically respond to a variety of texts/genres. Connect ideas in the titles and in the sequence of events across the texts. Identify common grammatical conventions in the use of verb forms and in spelling of some frequently used two-syllable words. Structure a story that has a beginning, a complication and conclusion. Draw additional details beyond the prompts.
4 *Expected level for Year 4 462.5 to < 487.5	Locate directly stated information in a variety of genres. Recognise the correct grammatical conventions in the use of capitals for proper nouns and in spelling of blends. Write a coherent text that has a few simple ideas by using common story elements, such as a simple title, and has a beginning but the conclusion may be missing or weak.
3 437.5 to < 462.5	Locate the main events in a variety of texts. Identify common language conventions in the use of text connectives and synonyms. Spell diagraphs; identify and correct errors in some frequently used one-syllable words.
2 412.5 to < 437.5	Make some meaning from texts that have visual images. Identify setting, author and simple literal information explicitly stated in a variety of texts/genres. Demonstrate basic and emerging proficiency in the use of prepositions and pronouns. Write a text consisting of a few simple ideas but with a weak structure.
1 362.5 to < 412.5	Identify literal information that is directly stated such as the titles and important dates in a variety of texts/genres. Identify meanings of simple words used in context. Write ideas using simple vocabulary but structure is limited to one paragraph.
0 Less than 362.5	Students at this level are not able to do any of the skills above and/or there is insufficient evidence to indicate their ability.

2.5.4.3 DEVELOPMENT OF THE EXPECTED LEVELS AND BENCHMARKS IN LITERACY AND NUMERACY

The purpose of the expected level in literacy and numeracy is to provide a reference point for the countries to indicate the minimum standard of achievement for students who have gone through four and six years of schooling. It is also to provide countries with information about how their students have performed in relation to the expected level.

The reference points were derived from the set of learning outcomes indicated on the eight-level proficiency scale (Tables 2.4 and 2.5 above), which was developed using the psychometric analysis of the 2012 and 2015 tests. This scale shows what students are able to do to qualify for each performance level. These learning outcomes are based on the regional benchmark indicators. Subject experts in English and mathematics at EQAP, who were involved in developing the learning descriptors and item construction, were tasked to set the reference points for literacy and numeracy for years 4 and 6.

The process of setting the expected levels entailed discussing the learning outcomes on the proficiency scale, focusing on the specific skills and knowledge that are represented at each level of the scale. The expected levels were then finalised, based on how the learning outcomes mapped the regional benchmark indicators in literacy and numeracy.

The expected level 4 and level 5 were finalised as the benchmarks for years 4 and 6 respectively for literacy. The same process was used for numeracy, where level 3 and level 5 were finalised as the benchmarks for years 4 and 6, respectively.

2.6 LIMITATIONS/CHALLENGES

As noted in the introductory section of this report, PILNA 2012 was designed initially as a one-time snapshot of literacy and numeracy levels in the region with the goal of gaining insight into student learning outcomes in literacy and numeracy in Year 4 and Year 6. In 2014, however, FEdMM elected to implement a second cycle of PILNA in 2015. The challenge was the relatively short period of time in which to develop improvements in design, item development, administration and data analysis. For this reason, PILNA 2015 was implemented with a view to strengthening elements of the administration for future cycles.

A partnership between EQAP and ACER was established to support particular areas of technical expertise, including psychometric support, sampling support, questionnaire support, and support for trend measures and reporting. This partnership was developed to support the long-term strengthening of the PILNA programme.

The results in this report, and any differences reported between student groups, are raw results and differences, unadjusted for any relevant background factors that might explain the reported observations. This fact applies to all comparisons in this report. Further analysis could usefully be conducted to identify relevant contributing factors, and until that is done, care should be taken in the way the results are interpreted.

The questionnaire component of PILNA was developed as a pilot in 2015. While the findings of the pilot have resulted in refinement of the student, teacher and head teacher questionnaires, it is not possible to report extensively on the questionnaire responses. Responses are included only as a demonstration of the possibility of using contextual data in future PILNA cycles.

There was also the implementation of a hybrid coding scheme described earlier in this chapter. Only a select number of items were able to be coded, but the longer-term plan is to develop assessments that are implemented using a full coding scheme. As with the questionnaires, while not a limitation of 2015, the hybrid coding – and subsequent planned coding study – will inform cognitive data collection and analysis for future cycles.

The categorisation of locality (urban, rural, remote or very remote) is not included in the regional and small islands states reports because of the differing definitions of locality in each of the PILNA countries. For example, one country identifies all its schools as rural. However, locality is included as a subgroup in the individual country reports.

3 CHAPTER 3: PERFORMANCE OF YEAR 4 AND YEAR 6 STUDENTS IN NUMERACY

This chapter describes the overall achievement of Year 4 and Year 6 students in the 2015 PILNA assessment of numeracy. The results are disaggregated by the three strands of numeracy (numbers, operations, and measurement and data), as well as by gender and school authority. This chapter also presents data on regional trends in overall numeracy performance between the PILNA cycles of 2012 and 2015.

3.1 GENERAL INFORMATION ON STUDENT NUMBERS

In total, 11,289 Year 4 students and 14,292 Year 6 students participated in the PILNA Numeracy test. One country, Papua New Guinea, required that only students with six years of formal education (equivalent Year 6) participate in PILNA. Table 3.1 shows the number of students disaggregated by group (gender and school authority) and by year level (Year 4 and Year 6). As can be seen in the table, similar numbers of boys and girls in both Year 4 and Year 6 participated in the PILNA numeracy assessment. In both year levels, a higher proportion of the participating students were from government schools; 87% of Year 4 students and 81% Year 6 students were from government schools, while 13% of Year 4 students and in 19% of Year 6 students were enrolled in non-government schools.

Table 3.1: Student numbers, PILNA 2015 numeracy assessment

Year	Year 4			Year 6		
Gender	Girls	Boys	**Total**	Girls	Boys	**Total**
	5655	5634	**11289**	7179	7113	**14292**
Authority Type	Non-Government	Government	**Total**	Non-Government	Government	**Total**
	1446	9843	**11289**	2715	11577	**14292**

3.2 OVERALL PERFORMANCE IN NUMERACY

3.3 COMPARING YEAR 4 AND YEAR 6

The data show very substantial improvement in numeracy achievement across the region as students progress from Year 4 to Year 6. This improvement is also evident in levels of numeracy at each year level between the 2012 and 2015 cycles of PILNA. (Proficiency level descriptors can be referred to in Table 2.4 of Chapter 2).

The overall improvement in numeracy achievement in terms of the distribution of students on the uniform proficiency scale from Year 4 to Year 6 can be seen Table 3.2 and in the corresponding histogram, Figure 3.1. The proportion of students in the three highest proficiency levels of the numeracy scale is markedly higher for Year 6 students than for Year 4 students. Similarly, the proportion of students performing at the lowest proficiency levels (Level 4 and below) is substantially lower for Year 6 students than for Year 4 students. These findings suggest improvement in learning outcomes in numeracy as students progress from Year 4 to Year 6.

Table 3.2: Distribution of students by year and proficiency levels, PILNA 2015

Proficiency level	Percentage	
	Year level	
	4	6
8	8.36	16.30
7	10.49	15.29
6	15.61	18.03
5	19.47	18.32
4	19.50	15.19
3	12.78	8.42
2	7.55	4.52
1	5.51	3.42
0	0.73	0.51

Expected proficiency level for Year 6

Expected proficiency level for Year 4

The corresponding histogram (Figure 3.1) displays the distribution of proficiency level achievement at Year 4 and Year 6 in 2015.

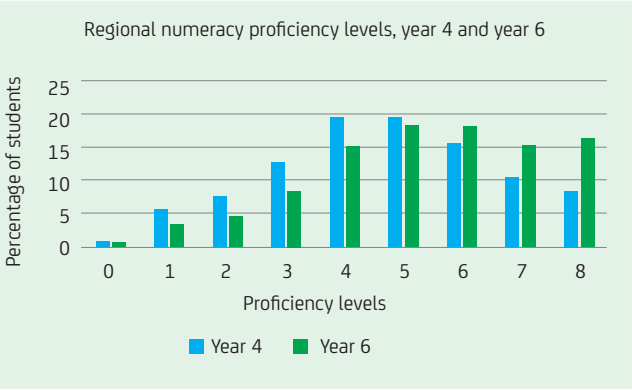


Figure 3.1: Regional numeracy proficiency levels, Year 4 and Year 6, PILNA 2015

The stacked graph (Figure 3.2) is another visual representation of Table 3.2, where the distribution of percentages represents students achieving at each proficiency level between Year 4 and Year 6.

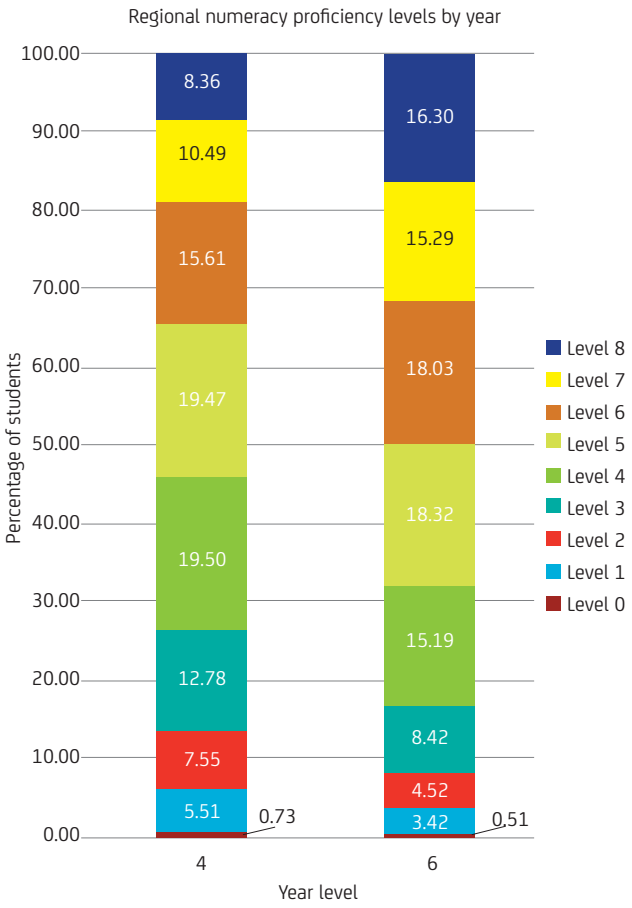


Figure 3.2: Regional proficiency levels in numeracy by year level, PILNA 2015

Figure 3.3 is a graphical representation of the distribution of scores in numeracy for Year 4 and Year 6 in 2015. The distribution is relatively symmetrical, with the range of scores in the upper half of the box approximately the same as the range of scores in the lower half of the box for both years. As described in Chapter 1, box plots have the advantage of comparing samples or subgroups at one time on a common scale, making differences in learning achievement readily apparent.

It can also be seen that the overall spread of scores, as well as the inter-quartile range, is similar in both Year 4 and Year 6. The distribution for Year 6 is wider than that for Year 4 students. It can also be seen that the boxes and whiskers of the box plot are about the same height for both Year 4 and Year 6, indicating a similar spread and distribution shape in the proficiency of both groups. Again, the distribution of scores indicates overall consistent growth in student learning outcomes between Year 4 and Year 6. In other words, on average, students are progressing in their growth between Year 4 and Year 6 at a more or less consistent rate across all ability levels.

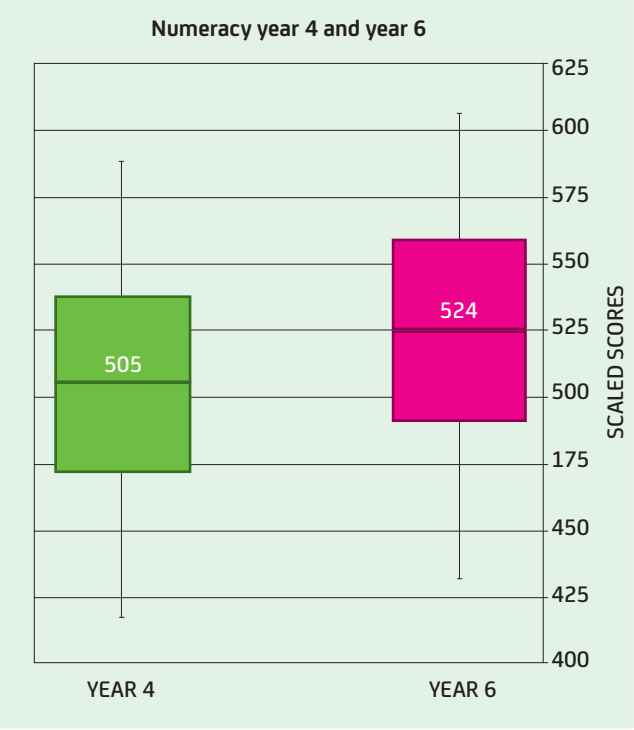


Figure 3.3: Distribution of numeracy scores for Year 4 and Year 6, PILNA 2015

3.4 COMPARING 2012 AND 2015

Comparing the 2012 and 2015 cycles of PILNA, there is notable improvement across the region in numeracy achievement in terms of the distribution of student percentages in the eight proficiency levels. This distribution is presented in Table 3.3. Fifty-four per cent of 2015 Year 4 students were in the four highest proficiency levels, compared to 40% in 2012. Similarly, at Year 6 almost 50% of students were in the three highest proficiency levels in 2015, compared to 36% in 2012.

Proficiency level	Percentage			
	Year level			
	4		6	
	2012	2015	2012	2015
8	7.59	8.36	10.82	16.30
7	7.11	10.49	10.80	15.29
6	9.65	15.61	14.78	18.03
5	16.12	19.47	20.26	18.32
4	17.49	19.50	16.74	15.19
3	16.27	12.78	11.17	8.42
2	11.07	7.55	6.65	4.52
1	10.80	5.51	6.30	3.42
0	3.91	0.73	2.48	0.51

Expected proficiency level for Year 6

Expected proficiency level for Year 4

This improvement suggests strong growth in student learning in numeracy between 2012 and 2015, as depicted in Figures 3.4 and 3.5. Seventy-four per cent of Year 4 students were at or above the expected proficiency level (level 3 and above) in 2012, and this increased to 86% at or above the expected level proficiency level in 2015.

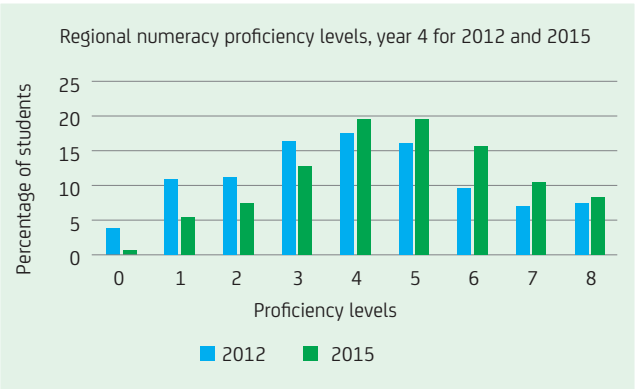


Figure 3.4: Regional numeracy proficiency levels, Year 4, PILNA 2012 and 2015

Similarly, the proportion of Year 6 students at or above the expected level (Level 5 and above) increased from a little over 56% in 2012 to almost 68% of students at or above the expected proficiency level in 2015. This improvement trend is shown in Figure 3.5.

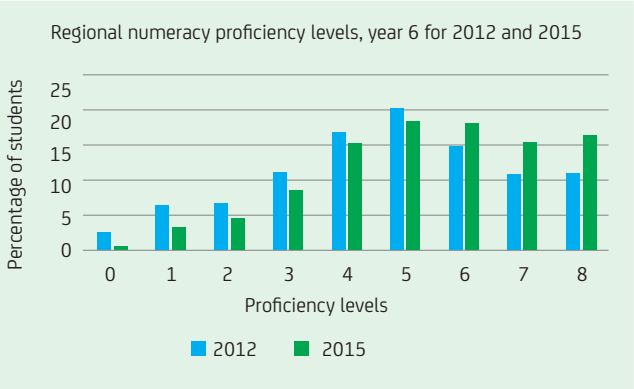


Figure 3.5: Regional numeracy proficiency levels, Year 6, for 2012 and 2015

3.5 YEAR 4 AND YEAR 6 DOMAIN AND STRANDS PERFORMANCE IN NUMERACY

The average regional improvement in numeracy is seen in all strands of the domain – numbers, operations, and measurement and data. The improvements are of similar magnitude as those in the general domain of numeracy at both year levels between 2012 and 2015. In other words, overall numeracy performance improved from 2012 to 2015 for both Years 4 and 6.

A comparison of mean performance by strand is presented in Table 3.4. In 2015, the highest mean performance of the Year 4 students was in measurement and data, while the highest mean performance of Year 6 students was in numbers. The increases in mean performance in numeracy between 2012 and 2015 are also presented in Table 3.4.

Table 3.4: Performance in numeracy by strand and overall trend, PILNA 2012 and 2015

Year	Descriptive statistics	Domain	Strands 2015			
		Numeracy 2012	Numeracy 2015	Numbers	Operations	Measurement and Data
4	Mean	486.43	505.01	504.01	505.29	508.03
	SD	61.82	51.35	66.74	50.94	48.54
6	Mean	506.01	523.47	528.16	522.83	522.88
	SD	58.95	53.22	68.04	53.45	50.29

SD=Standard Deviation

3.6 YEAR 4 AND YEAR 6 NUMERACY PERFORMANCE BY GENDER

Girls demonstrated higher levels of numeracy than boys, on average, across the region in 2015. This is evident in their higher mean scores as well as in the way students are distributed across the proficiency levels, with girls having a slightly higher proportion of students in the upper proficiency levels than boys in both Year 4 and Year 6. (See Table 3.5)

Table 3.5: Distribution of students by gender, year level and proficiency level, PILNA 2015

Proficiency level	Percentage			
	Year level			
	4		6	
	Girls	Boys	Girls	Boys
8	8.99	7.73	17.26	15.34
7	11.24	9.73	16.32	14.24
6	16.61	14.60	18.58	17.48
5	20.29	18.64	18.21	18.43
4	19.30	19.71	14.69	15.69
3	11.74	13.82	7.83	9.01
2	6.81	8.30	3.89	5.17
1	4.45	6.58	2.84	4.01
0	0.56	0.89	0.38	0.64

Expected proficiency level for Year 6

Expected proficiency level for Year 4

More girls than boys performed above the expected proficiency level at both Year 4 and Year 6. Boys outnumbered girls in the lower proficiency levels – at each level from level 4 and below at Year 4, and at each level from level 5 and below for Year 6. These distributions by gender and proficiency level are shown for Year 4 in Figure 3.6 and Year 6 in Figure 3.7.

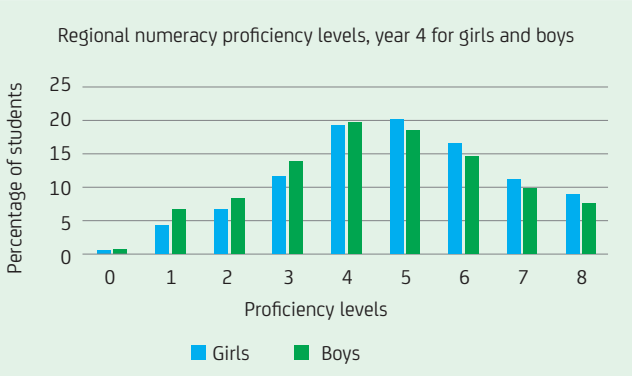


Figure 3.6: Regional numeracy proficiency levels, Year 4 girls and boys, PILNA 2015

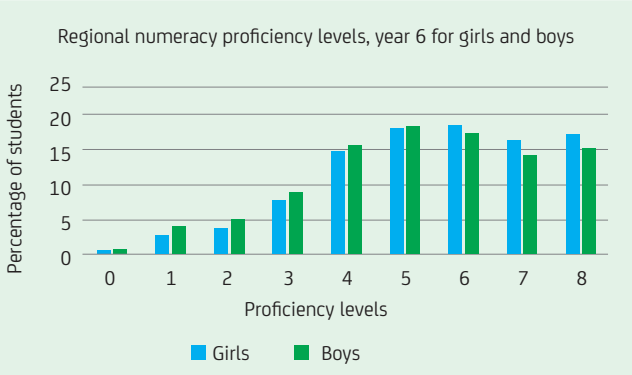


Figure 3.7: Regional numeracy proficiency levels, Year 6 for girls and boys, PILNA 2015

Table 3.6 indicates that girls outperformed boys in the overall numeracy domain, as well as in each strand in Year 4 and Year 6. However, the difference between the mean performance of girls and that of boys is relatively small.

The box plot in Figure 3.8 shows the distribution of scores in numeracy for Year 4 and Year 6 grouped by gender. The distribution is relatively symmetrical, with the range of scores in the upper half approximately the same as the range of scores in the lower half. However, the overall range is slightly wider for boys than for girls, i.e. the distribution of scores is more widely dispersed among boys than among girls. Year 6 boys performed at a range of levels, encompassing the range of scores for girls in Year 4.

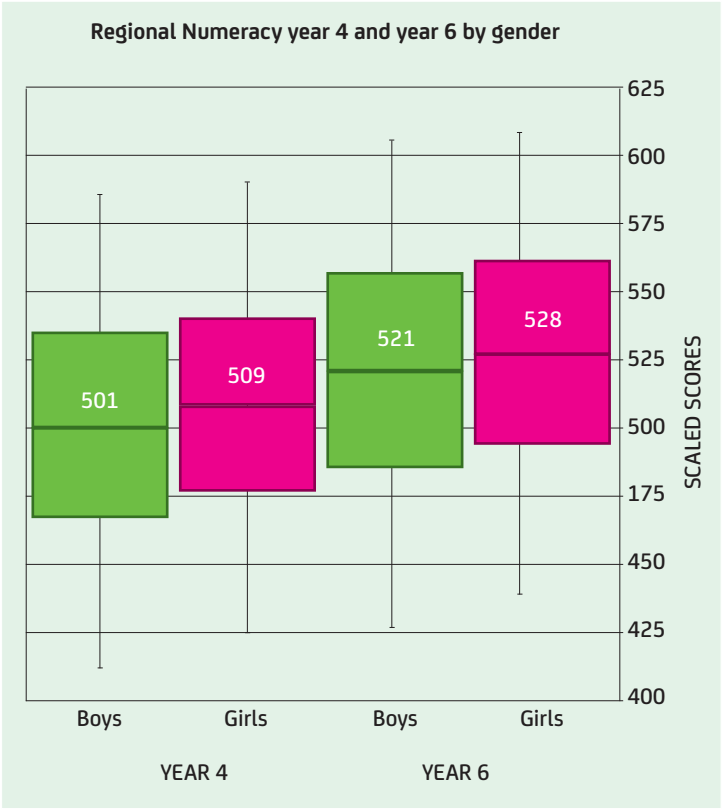


Figure 3.8: Distribution of numeracy scores by gender for Year 4 and Year 6, PILNA 2015

Table 3.6: Performance by gender in overall numeracy and by strand, PILNA 2015

YEAR	Gender	Descriptive statistics	Domain	Strands		
			Numeracy	Numbers	Operations	Measurements and Data
4	Female	Mean	508.70	508.49	509.05	510.72
		SD	50.12	65.04	49.82	47.72
	Male	Mean	501.31	499.52	501.51	505.34
		SD	52.30	68.11	51.76	49.21
6	Female	Mean	526.67	531.37	526.39	524.64
		SD	51.97	66.46	52.38	49.18
	Male	Mean	520.23	524.92	519.24	521.09
		SD	54.27	69.46	54.27	51.33

SD=Standard Deviation

3.7 YEAR 4 AND YEAR 6 NUMERACY PERFORMANCE BY SCHOOL AUTHORITY

In 2015, there was a slight difference in the distribution of government and non-government school students across the proficiency levels in numeracy. However, the difference in distribution was not uniform across year levels.

In Year 4, non-government schools have a slightly higher proportion of students (22%) in the upper two proficiency levels compared to students from government schools (18%). This difference is reversed in the top two highest proficiency levels in Year 6, where there are proportionally more government school students (33%) than non-government school students (25%). Table 3.7 presents the proportion of distribution across proficiency levels by year level and school authority.

Table 3.7: Distribution of students by school authority, year level and proficiency level, PILNA 2015

Proficiency level	Percentage			
	Year level			
	4		6	
	Non-government	Government	Non-government	Government
8	10.79	8.02	12.40	17.22
7	11.58	10.31	12.63	15.87
6	16.36	15.47	18.88	17.80
5	20.99	19.21	21.73	17.53
4	19.76	19.48	18.53	14.42
3	11.15	13.05	9.40	8.21
2	5.75	7.82	3.83	4.70
1	3.38	5.84	2.20	3.72
0	0.24	0.80	0.40	0.54

Expected proficiency level for Year 6

Expected proficiency level for Year 4

The comparative mean performance of government and non-government schools varies between year levels. Generally, in Year 4, students from non-government schools slightly outperformed students from government schools, based on the average regional achievement in numeracy, as seen in Figure 3.9.

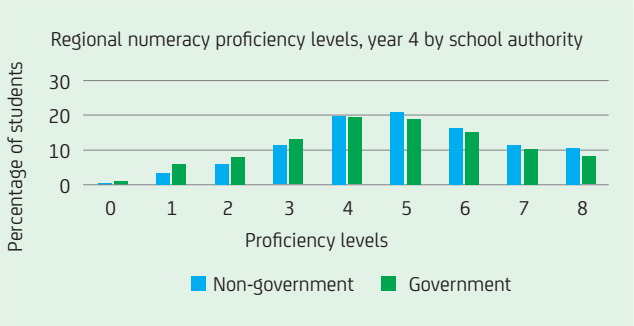


Figure 3.9: Regional numeracy proficiency levels, Year 4, by school authority type, PILNA 2015

Figure 3.10 shows the regional numeracy proficiency levels for Year 6 students by school authority type. It can be seen that at this level the students from government schools slightly outperformed those from non-government schools. However, the difference in mean performance between school authority types is small.

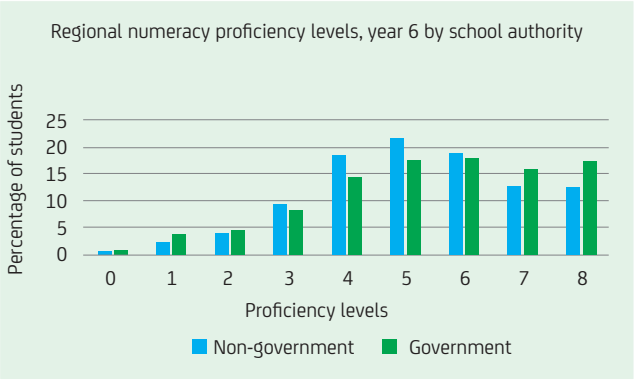


Figure 3.10: Regional numeracy proficiency levels, Year 6, by school authority type, PILNA 2015

Table 3.8 shows that Year 4 students in non-government schools outperformed students in government schools in each strand, as well as in the overall numeracy domain. And, as with the numeracy domain, the reverse situation is evident in Year 6; students from government schools slightly outperformed students from non-government schools in each strand. However, the difference in mean performance between government and non-government schools is relatively small.

Table 3.8: Performance in overall numeracy and strands by year level and school authority, PILNA 2015

YEAR	Authority type	Descriptive statistics	Domain	Strands		
			Numeracy	Numbers	Operations	Measurements and Data
4	Non-Government	Mean	512.99	514.57	512.90	515.40
		SD	49.07	64.86	49.06	46.84
	Government	Mean	503.80	502.43	504.12	506.94
		SD	51.60	66.92	51.15	48.73
6	Non-Government	Mean	519.50	525.22	518.00	520.90
		SD	48.49	61.88	48.94	45.95
	Government	Mean	524.34	528.76	523.90	523.30
		SD	54.28	69.43	54.43	51.28

SD=Standard Deviation

The box plot in Figure 3.11 shows the distribution of scores in numeracy for Year 4 and Year 6 grouped by school authority type. The distribution for each group is relatively symmetrical, with the range of scores in the upper half approximately the same as the range of scores in the lower half.

While the top Year 6 students in government schools performed better than the top Year 6 students in non-government schools, it can also be observed that the overall spread and inter-quartile range are wider for the government school students in Year 6 than the other groups. In other words, in Year 6, the distribution of scores among government school students is more widely dispersed than it is among the non-government school students.

This finding suggests that Year 6 students in government schools are performing at a wide range of levels. If the distribution were narrower, students, particularly at the lower proficiency levels, could be meeting expected learning outcomes.

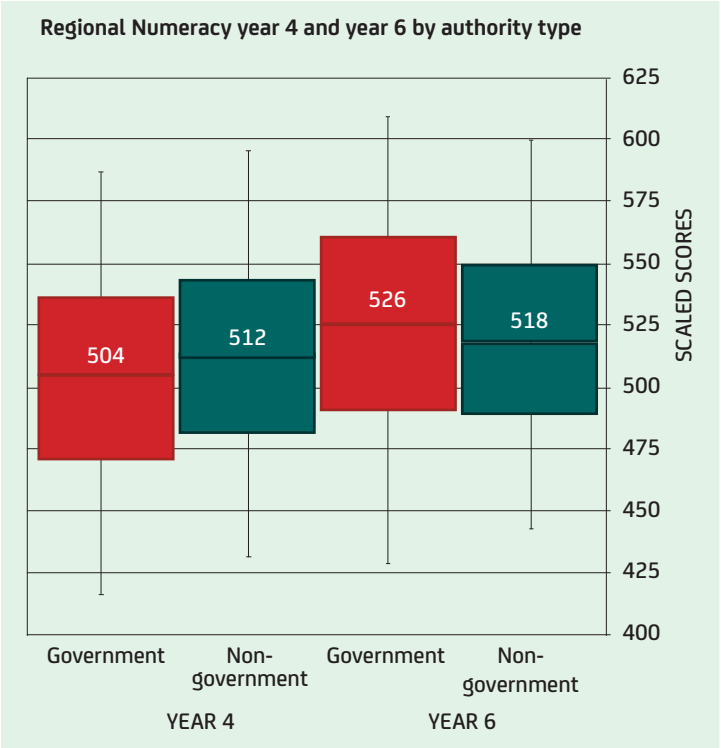


Figure 3.11: Distribution of numeracy scores by school authority for Year 4 and Year 6, PILNA 2015

3.8 CONCLUSIONS

The main findings discussed in this chapter are that overall achievement in numeracy has markedly improved across the region. Larger proportions of students at both Year 4 and Year 6 are reaching the higher proficiency levels. Larger proportions of students at both year levels are at or above the expected proficiency levels. And, as would be expected, substantial improvement is observed as students progress from Year 4 to Year 6.

Improvement in numeracy performance was observed between 2012 and 2015 and across both Year 4 and Year 6. Students were assessed for numeracy skills in the strands of numbers, operations, and measurement and data. In addition to improved performance in the general domain of numeracy, student mean performance also improved in all strands of the domain. Year 4 students had the highest performance in measurement and data, while Year 6 students had the highest performance in numbers.

At the regional level, girls outperformed boys in numeracy in both Year 4 and Year 6, although the difference in performance between boys and girls is minimal. Girls also performed better than boys at both year levels in all the strands. There is a slight difference in the distribution of boys and girls across the proficiency levels. Girls represent a slightly higher proportion of students in the upper proficiency levels than boys in both Year 4 and Year 6.

Finally, the comparative mean performance of students in government and non-government schools varies. In Year 4, students from non-government schools slightly outperformed students from government schools, based on the average regional achievement in numeracy. Year 4 students in non-government schools also outperformed students in government schools in the overall numeracy domain, as well as in all the strands. However, Year 6 students from government schools slightly outperformed students from non-government schools.

The following chapter, Chapter 4, discusses Year 4 and Year 6 students' performance in literacy for PILNA 2015.

4 CHAPTER 4: PERFORMANCE OF YEAR 4 AND YEAR 6 STUDENTS IN LITERACY

This chapter describes the overall achievement of Year 4 and Year 6 students in the 2015 PILNA assessment of literacy. The results are disaggregated by the three strands of literacy (reading, language features and writing), as well as by gender and school authority. This chapter also presents data on the regional trends in overall literacy performance between the PILNA cycles of 2012 and 2015.

4.1 GENERAL INFORMATION ON STUDENT NUMBERS

In total, 12,194 Year 4 students and 15,807 Year 6 students participated in the PILNA literacy test. One country, Papua New Guinea, required that only students with six years of formal education (equivalent Year 6) participate in PILNA. Table 4.1 shows the number of students disaggregated by group (gender and school authority) and by year level (Year 4 and Year 6). As can be seen in the table, similar numbers of boys and girls in both Year 4 and Year 6 participated in the PILNA literacy assessment. In both year levels, a higher proportion of the participating students were from government schools; 88% of Year 4 students and 79% Year 6 students were from government schools, while 12% of Year 4 students and 21% of Year 6 students were enrolled in non-government schools.
Table 4.1: Student numbers, literacy 2015 PILNA

4.2 OVERALL PERFORMANCE IN LITERACY

4.3 COMPARING YEAR 4 AND YEAR 6

The data discussed in this chapter show that across the region there has been some improvement in literacy achievement in terms of distribution of students across the proficiency levels between the 2012 and 2015 cycles of PILNA. However, the improvement has been inconsistent.

Looking at the 2015 regional literacy performance of students as they progress from Year 4 to Year 6 there was an improvement, with a smaller percentage of students in the lower proficiency levels in Year 6 than in Year 4, and a bigger percentage of students in the higher proficiency levels (levels 7 and 8) in Year 6 than in Year 4. This improvement was, however, minimal, as indicated in Table 4.2 and Figures 4.1 and 4.2, which show the comparative distribution of students for each proficiency level. Moreover, the proportion of students at or above the expected proficiency level (the regional minimum benchmark indicator highlighted in Table 4.2) was 46% for both Year 4 students and Year 6 students. (Proficiency level descriptors can be referred to in Table 2.5 of Chapter 2.)

The majority of students in both Year 4 and Year 6 are clustered between levels 3 and 6. Just under 50% of students performed at or above the expected proficiency levels in 2015.
Table 4.2: Distribution of students by year and proficiency level, PILNA 2015

The proportion of students performing at or above the expected proficiency level (level 5) was higher for Year 6 students than for Year 4 students. The histogram in Figure 4.1 represents the distribution of proficiency level achievement in Year 4 and Year 6.

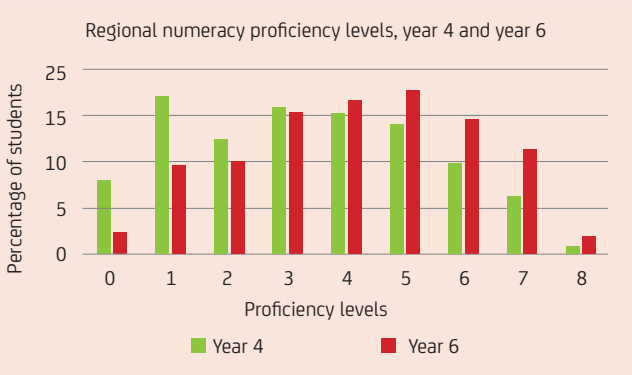


Figure 4.1: Regional literacy proficiency levels, Year 4 and Year 6, PILNA 2015

The stacked graph (Figure 4.2) is a visual representation of the distributions in Table 4.2, where the distribution of percentages shows an increase from Year 4 to Year 6.

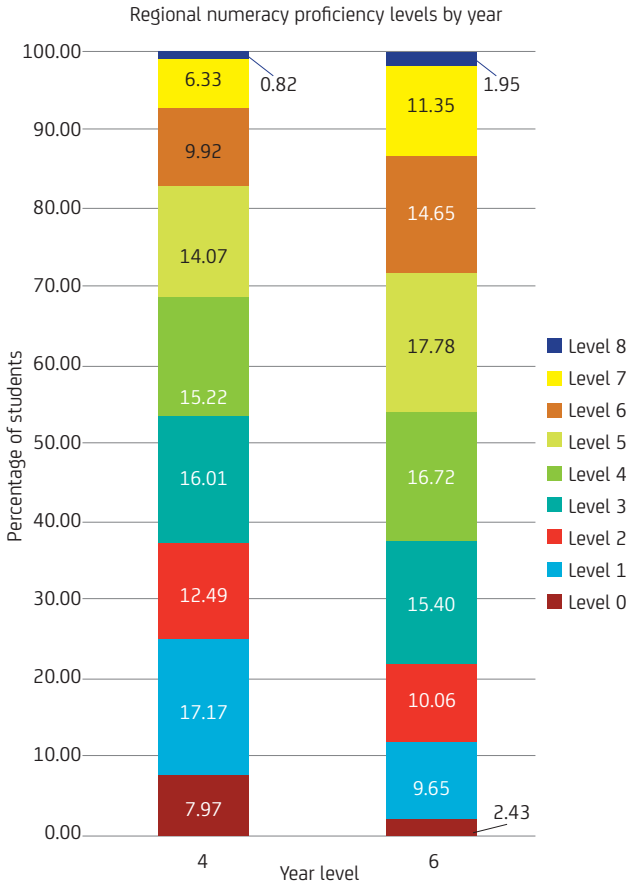


Figure 4.2: Regional proficiency levels in literacy by year level, PILNA 2015

The box plot (Figure 4.3) is a graphical representation of the distribution of scores in literacy for Year 4 and Year 6 in 2015. The distribution is relatively symmetrical, with the range of scores in the upper half approximately the same as the range of scores in the lower half for both years.

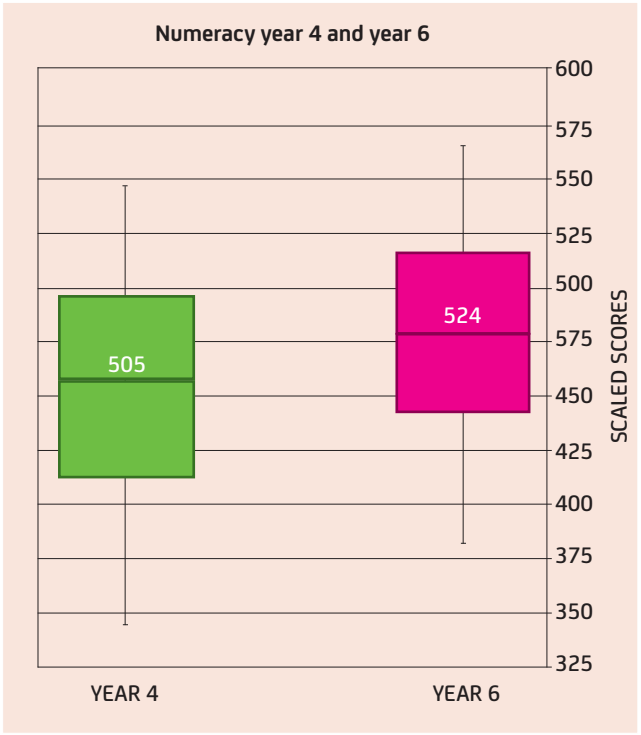


Figure 4.3: Distribution of literacy scores for Year 4 and Year 6, PILNA 2015

It can also be seen that the overall spread of scores, is similar in Year 4 to that of Year 6 in that the scores tend to bunch around the middle of the score range and tail off to a similar degree for students above and below the middle. The distribution for Year 6 is above that for Year 4, although the increase is slightly less in 2015 than in 2012. (In 2012 there was an increase in the literacy domain of just over 30 points, while in 2015 that increase was just under 25 points).

4.4 COMPARING 2012 AND 2015

Table 4.3 and Figures 4.4 and 4.5 show that the proportion of students in the lowest proficiency levels (levels 0 to 2 in Year 4, and Levels 0 and 1 in Year 6) has decreased since 2012. In 2012, 43% of Year 4 students were in the three lowest proficiency levels, compared to 38% in 2015; and, 16% of Year 6 students were in the two lowest proficiency levels in 2012, compared to only 12% in 2015.

There has, however, been a slight decline in the proportion of students in the uppermost two levels (levels 7 and 8), in both Years 4 and 6, between 2012 and 2015. In Year 4, 9% of students were in the two highest proficiency levels in 2012, compared to 7% in 2015 and, in Year 6, 20% of students were in the two highest proficiency levels in 2012, while there were only 13% in 2015.

Proficiency level	Percentage			
	Year level			
	4		6	
	2012	2015	2012	2015
8	1.59	0.82	5.24	1.95
7	7.59	6.33	14.98	11.35
6	9.15	9.92	13.44	14.65
5	12.14	14.07	14.73	17.78
4	12.82	15.22	12.59	16.72
3	14.13	16.01	12.81	15.40
2	13.51	12.49	9.99	10.06
1	18.03	17.17	10.55	9.65
0	11.05	7.97	5.68	2.43

Expected proficiency level for Year 6

Expected proficiency level for Year 4

Table 4.3: Distribution of students by year and proficiency levels between PILNA 2012 and 2015

Figures 4.4 and 4.5 are histograms representing the trends in the distribution of student achievement by proficiency level in literacy in Year 4 and Year 6.

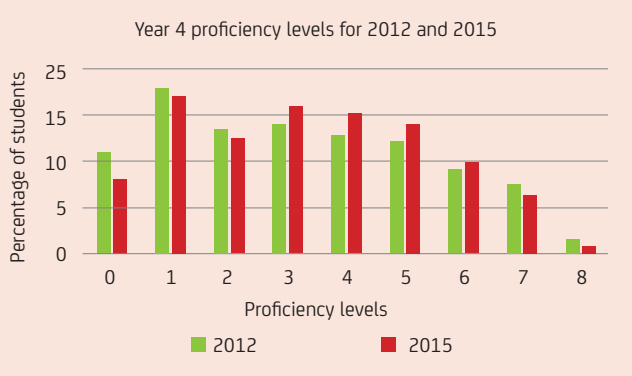


Figure 4.4: Regional literacy proficiency levels, Year 4, for PILNA 2012 and 2015

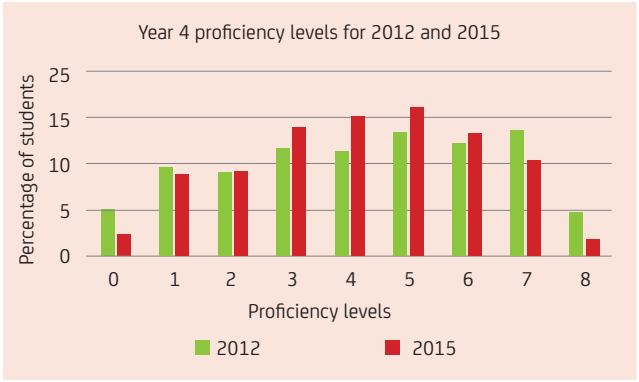


Figure 4.5: Regional literacy proficiency levels, Year 6, for PILNA 2012 and 2015

4.5 YEAR 4 AND YEAR 6 DOMAIN AND STRANDS PERFORMANCE IN LITERACY

Overall, the average regional achievement in literacy improved from Year 4 to Year 6. This increase can be seen in the overall literacy domain as well as in all the strands of the domain. The strands assessed for PILNA literacy were reading, language features and writing.

Table 4.4 shows that the Year 4 average mean performance in literacy increased between the 2012 and 2015 cycles of PILNA although, statistically, that for Year 6 revealed only flat, or minimal, change between 2012 and 2015.

Table 4.4 also shows that, among the strands, students in both Year 4 and Year 6 have the highest mean performance in reading, followed by language features, with writing having the comparatively lowest mean performance.

Table 4.4: Performance in literacy by strands and overall trends, PILNA 2012 and 2015

Year	Descriptive statistics	Literacy 2012	Literacy 2015	Strands 2015		
				Reading	Language Features	Writing
4	Mean	448.85	453.43	458.71	450.66	442.48
	SD	68.87	61.33	50.95	71.11	95.49
6	Mean	479.81	478.34	480.29	480.08	471.02
	SD	70.20	55.50	47.66	65.35	89.07

SD=Standard Deviation

4.6 YEAR 4 AND YEAR 6 LITERACY PERFORMANCE BY GENDER

Girls demonstrated higher levels of literacy than boys, on average, across the region in 2015. This is evident in their higher mean scores as well as in the way students are distributed across the proficiency levels, and is observed at both Year 4 and Year 6 (Table 4.5, Figures 4.6 and 4.7).

Table 4.5: Distribution of students by gender, year level and proficiency level, PILNA 2015

Proficiency level	Percentage			
	Year level			
	4		6	
	Girls	Boys	Girls	Boys
8	1.15	0.49	2.87	1.03
7	8.37	4.29	14.10	8.62
6	11.90	7.94	16.18	13.15
5	15.62	12.53	18.43	17.14
4	15.20	15.24	16.77	16.67
3	15.32	16.69	14.30	16.49
2	11.55	13.42	8.76	11.35
1	1v4.96	19.38	7.14	12.14
0	5.92	10.01	1.44	3.41

Expected proficiency level for Year 6

Expected proficiency level for Year 4

More girls than boys performed above the expected proficiency level for both Year 4 and Year 6. Boys outnumbered girls in the lower proficiency levels – at each level from level 4 and below at Year 4, and each level at level 5 and below for Year 6.

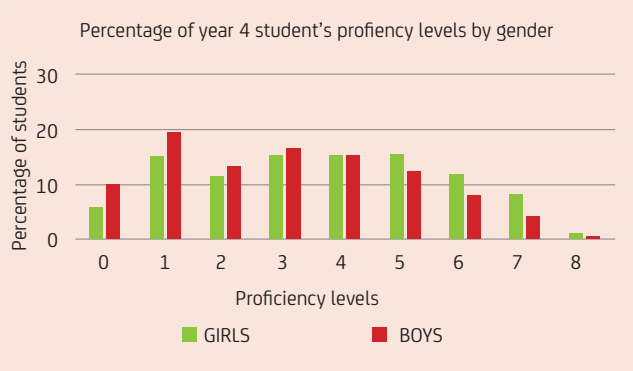


Figure 4.6: Regional literacy proficiency levels, Year 4 girls and boys, PILNA 2015

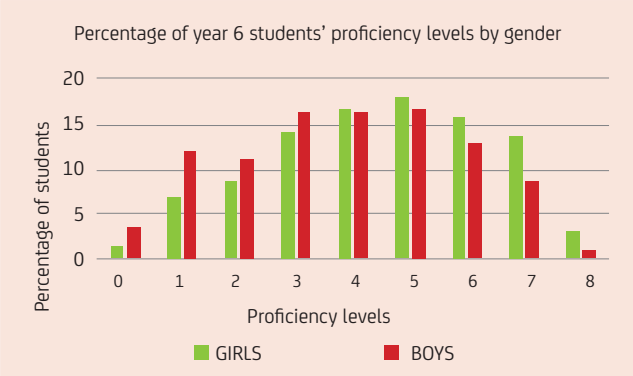


Figure 4.7: Regional literacy proficiency levels, Year 6 girls and boys, PILNA 2015

Table 4.6 shows that girls outperformed boys in the overall literacy domain, as well as in each strand in both Year 4 and Year 6.

The difference between girls and boys in overall literacy is larger than in numeracy, with girls approximately 13 points higher than boys in Year 4 and about 14 points higher than boys in Year 6. The difference is most pronounced in the writing strand, where Year 4 girls scored on average 22 points higher than boys, and at Year 6 about 27 points higher.
Table 4.6 shows that in writing, Year 4 girls have a similar mean performance to Year 6 boys, and scored on average 25 points more than Year 4 boys. Year 6 girls scored almost 29 points higher on average than Year 6 boys. In the overall literacy domain, girls in both year levels scored on average 18 points more than boys.

Table 4.6: Performance by gender in overall literacy and by strands, PILNA 2015

YEAR	Authority type	Descriptive statistics	Domain	Strands		
			Literacy	Reading	Language Features	Writing
4	Girls	Mean	462.47	465.08	460.02	455.23
		SD	60.77	50.59	70.51	94.66
	Boys	Mean	444.41	452.35	441.33	429.77
		SD	60.55	50.52	70.48	94.61
6	Girls	Mean	487.34	486.39	488.43	485.54
		SD	54.20	46.49	64.12	88.15
	Boys	Mean	469.44	474.26	471.81	456.68
		SD	55.34	48.04	48.04	87.64

SD=Standard Deviation

The box plot in Figure 4.8 shows the distribution of scores in literacy for Year 4 and Year 6 grouped by gender. The distribution is relatively symmetrical, with the range of scores in the upper half approximately the same as the range of scores in the lower half. This box plot shows differences in the distribution of achievement between boys and girls.

It can also be observed that the inter-quartile range is similar for girls and boys in both Year 4 and Year 6. However, the overall range is slightly wider for Year 4 boys and girls. Most Year 6 boys performed, on average, at a similar level to Year 4 girls. These differences in the spread of scores can also be seen in the values of the standard deviation of scores reported in Table 4.6. The largest variation in the spread of scores is seen in Year 4 and Year 6 writing, where Year 4 girls and boys have standard deviations of 94.66 and 94.61, respectively. In Year 6, girls have a standard deviation of 88.15 and boys have a standard deviation of 87.64.

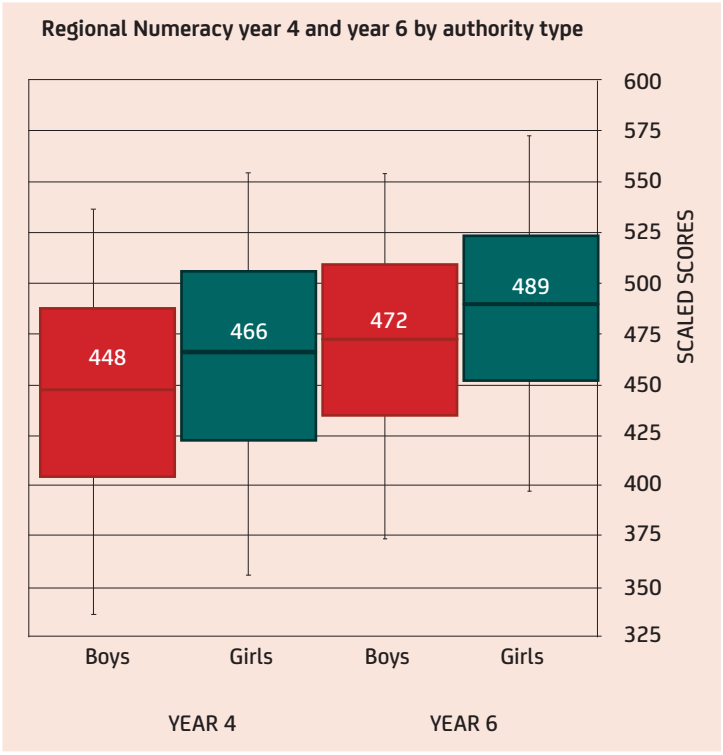


Figure 4.8: Distribution of literacy scores by gender for Year 4 and Year 6, PILNA 2015

4.7 YEAR 4 AND YEAR 6 LITERACY PERFORMANCE BY SCHOOL AUTHORITY

In 2015 there was a slight difference in the distribution of government and non-government school students across the proficiency levels in literacy. However, the difference in distribution is not uniform across year levels.

In both Year 4 and Year 6, non-government schools had a slightly higher proportion of students in the upper proficiency levels (levels 7 and 8) compared to students from government schools. And across all proficiency levels, students in non-government schools had a slightly higher mean performance in literacy than students in government schools. Table 4.7 presents the proportion of distribution across proficiency levels by year level and school authority.

Table 4.7: Distribution of students by school authority, year level and proficiency level, PILNA 2015

YEAR	Authority type	Descriptive statistics	Domain	Strands		
			Literacy	Reading	Language Features	Writing
4	Non-Government	Mean	459.37	467.34	460.79	441.48
		SD	63.84	53.22	73.32	97.32
	Government	Mean	452.69	457.57	449.32	442.73
		SD	60.85	50.49	70.66	95.16
6	Non-Government	Mean	488.10	490.44	493.99	479.85
		SD	52.89	44.49	60.90	84.09
	Government	Mean	475.71	477.57	476.36	468.61
		SD	55.90	48.11	65.99	90.24

SD=Standard Deviation

Table 4.7. The figures indicate the distribution of proficiency level achievement of Year 4 and Year 6 students by school authority.

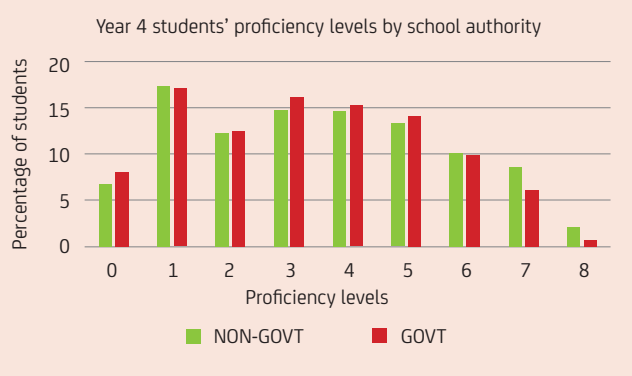


Figure 4.9: Regional literacy proficiency levels, Year 4, by school authority type, PILNA 2015

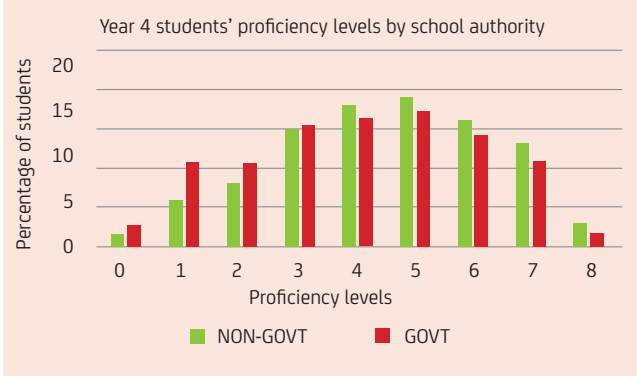


Figure 4.10: Regional literacy proficiency levels, Year 6, by school authority type, PILNA 2015

The performance in literacy in government and non-government schools shows a uniform increase from Year 4 to Year 6. However, the difference in both year levels is relatively small. These results can be observed in the domain as well as the strands, as evident in Table 4.8 and Figure 4.11.

Except for Year 6 students in non-government schools (who have the highest mean performance in language features), students consistently have the highest mean performance in the reading strand across both year levels and school authority, and the lowest mean performance in the writing strand. However, the difference in mean performance in the Year 4 writing strand between both school authorities and both year levels is negligible.

Table 4.8: Performance by school authority in overall literacy and by strands, PILNA 2015

YEAR	Authority type	Descriptive statistics	Domain	Strands		
			Literacy	Reading	Language Features	Writing
4	Non-Government	Mean	459.37	467.34	460.79	441.48
		SD	63.84	53.22	73.32	97.32
	Government	Mean	452.69	457.57	449.32	442.73
		SD	60.85	50.49	70.66	95.16
6	Non-Government	Mean	488.10	490.44	493.99	479.85
		SD	52.89	44.49	60.90	84.09
	Government	Mean	475.71	477.57	476.36	468.61
		SD	55.90	48.11	65.99	90.24

SD=Standard Deviation

The box plot in Figure 4.11 shows the distribution of scores in literacy for Year 4 and Year 6 grouped by school authority type. The distribution for each group is relatively symmetrical, with the range of scores in the upper half approximately the same as the range of scores in the lower half.

Year 4 students in government schools and non-government schools performed at similar levels in literacy, and it can also be observed that the overall range and inter-quartile range are wider for both school authority types in Year 4. In other words, the distribution of scores among Year 4 students from both types of schools is more widely dispersed than the distribution of scores among students from both types of school in Year 6. These results suggest that Year 4 students in government schools are performing at a wider range of levels in literacy.

These differences in the spread of scores can also be seen in the values of the standard deviation of scores reported in Table 4.8. The largest variation in the spread of scores is seen in Year 4 and Year 6 writing, where Year 4 students in non-government schools have a standard deviation of 97.32 and those in government schools have a standard deviation of 95.16. In Year 6, students in non-government schools have a standard deviation of 84.09 and government school students have a standard deviation of 90.24.

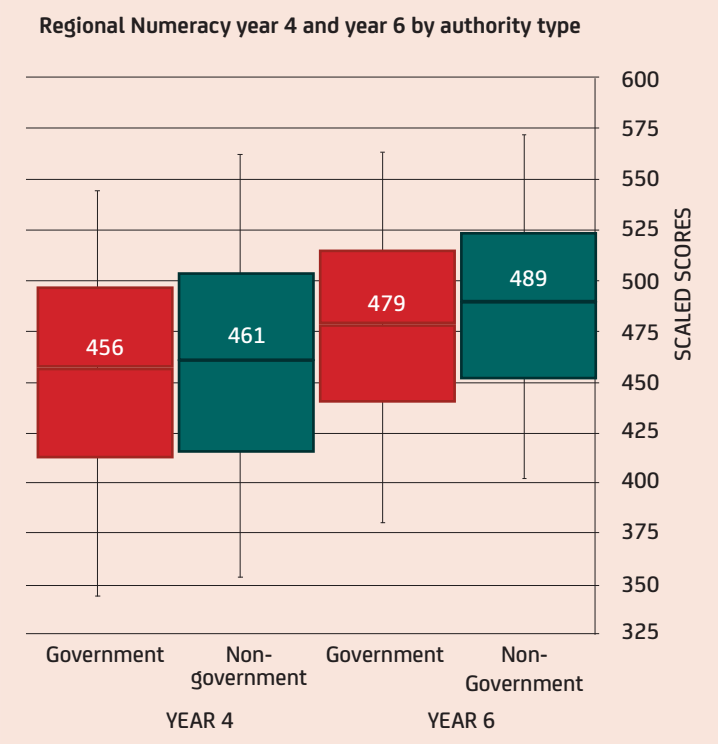


Figure 4.11: Distribution of literacy scores by school authority for Year 4 and Year 6, PILNA 2015

4.8 CONCLUSIONS

The main findings discussed in this chapter are that there has been some improvement in overall achievement in literacy across the region. Importantly, there has been some improvement in literacy achievement in terms of distribution of students across the proficiency levels between the 2012 and 2015 cycles of PILNA.

However, the improvement across proficiency levels is not consistent between 2012 and 2015. On the one hand, there are fewer students in the lowest proficiency levels and more students in the middle range (levels 3–6) of the proficiency scale in Years 4 and 6. At the same time, there has been a decline in the number of students achieving at the highest proficiency levels between 2012 and 2015. This shift toward the middle of the distribution signifies somewhat uneven progression in literacy between the two year levels.

Increases in achievement were evident in all the strands of the domain, as well as in the general domain of literacy from 2012 to 2015. The strands assessed for PILNA literacy were reading, language features and writing. The Year 4 average mean performance in literacy increased between the 2012 and 2015 cycles of PILNA. Statistically, however, the overall performance in Year 6 revealed flat, or minimal, change between 2012 and 2015.

Findings from PILNA 2015 showed a notable difference in the distribution of boys and girls across the proficiency levels in literacy. There were more girls than boys in the higher proficiency levels and more boys than girls in the lower proficiency levels in both years.

In each proficiency level at or below the expected level for Year 4 (level 4) and Year 6 (level 5), there are proportionally more boys than girls. At the same time, there are proportionally more girls than boys among those who are achieving in literacy above the expected proficiency level.

Finally, the performance in literacy in government and non-government schools shows a uniform increase from Year 4 to Year 6. Generally, students from non-government schools slightly outperformed students from government schools across all proficiency levels, although the difference was relatively small. And in both Year 4 and Year 6, non-government schools had a slightly higher proportion of students in the upper proficiency levels compared to students from government schools.

Overall, evidence from PILNA 2015 indicates that, at the regional level, achievement levels in numeracy and literacy for Year 4 and Year 6 have improved since 2012. Regional results show that there has been large improvement in performance in the numeracy assessment, while literacy results, though improved, have been more varied.

The following chapter discusses the coding processes implemented for PILNA 2015, and how the use of a full coding system for all assessment items in future cycles of PILNA will provide more specific knowledge about students' literacy and numeracy skills.

5 CHAPTER 5: CODING THE COGNITIVE INSTRUMENT IN PILNA 2015

A coding approach enables a process where additional information about student responses are captured rather than just scoring responses as 'correct-incorrect'. It was suggested that EQAP consider a conceptual shift from 'scoring' to 'coding' for future cycles; and, subsequently, selected items were coded during the data capture process.

While the scoring protocols for the 2015 cycle had already been established, a coding method of improving data capture for the future and testing improvements to data capture were trialled during the 2015 administration. A subset of the 2015 cycle's data was used to develop a PILNA coding scheme. This chapter discusses the development and use of a coding scheme for PILNA 2015. It also includes examples of how a coding scheme can be used for more in-depth analysis of student responses.

5.1 ADVANTAGES OF THE CODING SCHEME

In 2015, a coding scheme was piloted for certain items in the literacy and numeracy PILNA assessment instruments. The advantage of this coding scheme is that it enables a full range of responses from the assessed students to be analysed. This process provides information about why some incorrect responses are more frequently provided by students than others; which, in turn has the potential to indicate different levels of understanding or ability in relation to the concepts and skills underpinning a question.

Information collected about student responses can be shared with classroom teachers who can use the data to address misconceptions by students on specific topics. Item analysis based on capturing coded responses, therefore, presents an improved opportunity for teaching intervention.

5.2 THE CODING PROCESS

A review of the outcomes of PILNA 2012 suggested that more information about student understanding and misunderstanding could be obtained with a relatively minor adjustment to the item design and with only a small additional amount of work for those responsible for processing student responses.

For the 2015 administration, some items were scored using the same approach applied in the PILNA 2012 administration (i.e. only correct-incorrect). However, for selected items an additional type of incorrectness was included in the coding schedule and using additional numbered codes in order to differentiate possible misconceptions for particular content areas (see Figure 5.1). A number of items were readily amenable to this shift in data capture. This hybrid approach to coding was intended to make the transition to coding easier for the in-country staff doing the scoring/coding.

Coders were trained to recognise those different responses in the students' completed test forms and assign each observed response to one of the defined coding categories. This kind of analysis provides better estimates of student ability, as well as more detailed information and insight about learning outcomes. Responses from this pilot will be used to develop a feasible coding scheme for all the items in subsequent administrations.[12]

12 Because most coders were used to marking a '1' for correct answers, the coding scheme was adjusted so that the correct response was "force-coded" as '1' for some items where correctness was readily apparent – while other responses were also coded, rather than scoring '0', as demonstrated in Figure 5.1.

ITEM	STRAND	2015 DESCRIPTOR	RESPONSES	CODE	COMMENTS
10	NUMBERS	Round off numbers to the nearest hundredths(or to two decimal places, with the first digit > 5)	5.70 5.700 Other response No response	1 8 0 9	
14	NUMBERS	Compare decimal numbers with same numbers of decimal places and whole number > 50	= < > Other response No response	1 2 3 0 9	
24	OPERATIONS	Add proper fractions with 1 or 2 digit denominations that are multiples	11/10;1 1/10;1.1; 55/50 Common denominator (or its multiple) identified 9/15;3/5 Other response No response	1 2 8 0 9	"Straight addition"

Figure 5.1: Sample coding scheme

Figure 5.1 is an example of a coding scheme for numeracy. The "response" and "code" columns specify the code that is assigned to a particular response. For example, in Figure 5.1 for Item 14 (numeracy), a response of <, =, and > are assigned a code of 1, 2, and 3, respectively, while all other responses are assigned a code of 0; and no response is assigned a code of 9.

5.3 Examples of coding in numeracy and literacy

Tables 5.1 and 5.2 present example items that show possible misconceptions in the numeracy (Table 5.1) and literacy (Table 5.2) assessments. For each item, there is a 'notes and interpretation' section discussing the misconception and the response categories. The item analysis follows the example and described misconception. The analysis includes the scores assigned to each category, the descriptive statistic, and the average abilities of students who responded in each category.

The notes accompanying the literacy examples also include some suggested classroom activities designed to address the issues identified in the analyses.

Each analysis includes the following five elements:

1. 'Label' shows the different response categories that have been defined for the item (and explained in the 'notes and interpretation' column).
2. 'Score' shows the score assigned to each defined response category.
3. 'Count' and '% of total' show the number and proportion of students responding in each category.
4. 'Pt Bis' (point-biserial) is a calculated statistic that is the correlation between the ability of students responding in each category for this item and their ability as shown by the whole test. It indicates the degree of consistency of this item with the rest of the test. This is typically a positive number for correct responses, and is smaller or negative for partially correct, incorrect, or missing responses.
5. 'PV1Avg:1' (average ability of students in the category, calculated from a single plausible value) is a calculated statistic that indicates the average ability of students responding in each category. It can often show that the different response categories separate students into clearly distinct ability groupings.

5.3.1 NUMERACY

In the four examples following, the item analyses are provided in relation to a small selection of numeracy test questions. The Notes and interpretation column explains different aspects of understanding, misunderstanding, errors or lack of understanding that might be indicated by different student responses. The technical item analysis follows the discussion of student misconceptions.

Table 5.1: Examples of the analysis for numeracy items.

<table>
<tr><th></th><th colspan="6">ITEM AND RESPONSE DATA</th><th>NOTES AND INTERPRETATION</th></tr>
<tr><td></td><td colspan="6">Item:19:
Subtract

731
- 15</td><td>The correct answer is 716. In subtraction a common misconception is that students subtract the smaller number from the larger number irrespective of the position of the digits. So a common wrong answer is 724. In this case, the student subtracted 1 from 5 because 1 is smaller than 5. Here the assumption is that subtraction is commutative.</td></tr>
<tr><td></td><td colspan="6">Common Responses
716 Student correctly subtracts 15 from 731

726 or similar – Student correctly subtracts the ones digit (11 – 5 = 6) but does not complete the ‘borrowing’ correctly in the tens digit (3-1 = 2 instead of 2-1 = 1)

724 Student subtracts 1 from 5 to get 4 in the ones place instead of ‘borrowing’ to subtract 5 from 11, resulting in 6 in the ones place.</td><td>Intervention Idea: To emphasise the importance of order in subtraction (the concept that subtraction is not commutative) teachers might demonstrate the use of a method of subtraction such as ‘decomposition’ or ‘equal addition’.</td></tr>
<tr><td>Label</td><td>Response</td><td>Score</td><td>Count</td><td>% of tot</td><td>Pt Bis</td><td>PV1Avg:1</td><td rowspan="4">For this item, the average ability (PV1Avg:1) of those whose responses indicated the misconception (8) is lower than those who responded with a partially correct (2) answer, but above those getting the item incorrect (0). It seems to be a clearly identified group.</td></tr>
<tr><td>0</td><td>Any other response</td><td>0</td><td>3660</td><td>17.06</td><td>-0.46</td><td>-0.705</td></tr>
<tr><td>1</td><td>716 (correct)</td><td>2</td><td>12290</td><td>57.29</td><td>0.52</td><td>0.894</td></tr>
<tr><td>2</td><td>At least 1 column of correct subtraction</td><td>1</td><td>4648</td><td>21.66</td><td>-0.19</td><td>-0.014</td></tr>
<tr><td>8</td><td>724</td><td>0</td><td>856</td><td>3.99</td><td>-0.03</td><td>-0.188</td><td></td></tr>
<tr><td></td><td colspan="6">Scoring for this item allowed full credit for the correct answer and partial credit for responses that showed at least one column of correct subtraction. No credit was given for any other response. In this case, we see that over half of the students had the correct answer (57%) and an additional 22% of students could get part of the subtraction complete.</td><td></td></tr>
<tr><td></td><td colspan="6">Item:43:
Round this number to the nearest hundredth 8.303</td><td>A common misconception is that students confuse hundreds (3 digit number) with hundredths (2 decimal places). The students apply their understanding of hundreds having three digits to hundredths and round off leaving three numbers after the decimal point. When working with decimal numbers, the first digit after the decimal point is the tenth and the second is the hundredth. Rounding to the nearest hundredth should have two digits after the decimal point.</td></tr>
</table>

	ITEM AND RESPONSE DATA						NOTES AND INTERPRETATION
	Common Responses 8.30 Student correctly rounds the 3 in the thousands place down leaving 0 in the hundredths place 8.300 Student correctly rounds the 3 in the thousands place down but includes a zero in the thousands place rounding the number to the nearest thousandth rather than hundredth						
Label	**Response**	**Score**	**Count**	**% of tot**	**Pt Bis**	**PV1Avg:1**	For this item, the average ability (PV1Avg:1) of those whose responses indicated a misconception (8) is actually a little higher than those who responded with a correct answer. This indicates that knowledge of this piece of mathematics is not related to mathematical ability; rather, it is a clear case for focussed teaching of this concept.
0	Any other response	0	8728	74.01	-0.46	0.297	
1	8.30 (correct)	1	752	6.38	0.18	1.365	
8	8.300	0	2313	19.61	0.4	1.374	
	Scoring for this item allowed full credit for the correct response and no credit for any other response. In this case we see than only 6% of students were able to correctly round to the nearest hundredth while almost 20% were able to round the number but expressed it in terms of thousandths rather than hundredths. The data do not capture what the majority of students (74%) provided in response to the question but we know that some form of answer was provided. Blanks in every case were given code 9 to differentiate them from incorrect responses.						
	Item:56: 3 /4 + 1 /8 =						A common misconception is when adding fractions with different denominators; students simply add the numerators and the denominators without first finding the equivalent fractions with the same denominator. The same applies to subtraction of fractions with different denominators.
	Common Responses 7/8 Student converts 3/4 to 6/8 and correctly adds the numerators. Equivalent fractions are also correct i.e. 14/16, 21/24, 28/32 ? /8 Student correctly finds the common denominator but not the correct numerator 4 /12 Student adds 3 + 1 to get 4 and 4 + 8 to get 12						Intervention idea: Students might benefit from physical examples of fractions that they can manipulate to demonstrate how fractions are related to one another and to show the results of adding fractions.
Label	**Response**	**Score**	**Count**	**% of tot**	**Pt Bis**	**PV1Avg:1**	For this item, the average ability (PV1Avg:1) of those whose responses indicated a misconception (8) is a little below but close to those who responded with a partially correct answer (2). It seems to be an identifiable group, suggesting room to focus teaching on this particular error.
0	Any other response	0	5471	45.41	-0.47	0.021	
1	7/8 (correct)	2	1819	15.1	0.41	1.673	
2	Correct common denominator	1	1562	12.96	0.13	0.896	
8	4/12	0	3196	26.53	0.1	0.663	

	ITEM AND RESPONSE DATA						NOTES AND INTERPRETATION
	Common Responses 8.30 Student correctly rounds the 3 in the thousands place down leaving 0 in the hundredths place 8.300 Student correctly rounds the 3 in the thousands place down but includes a zero in the thousands place rounding the number to the nearest thousandth rather than hundredth						
Label	**Response**	**Score**	**Count**	**% of tot**	**Pt Bis**	**PV1Avg:1**	For this item, the average ability (PV1Avg:1) of those whose responses indicated a misconception (8) is actually a little higher than those who responded with a correct answer. This indicates that knowledge of this piece of mathematics is not related to mathematical ability; rather, it is a clear case for focussed teaching of this concept.
0	Any other response	0	8728	74.01	-0.46	0.297	
1	8.30 (correct)	1	752	6.38	0.18	1.365	
8	8.300	0	2313	19.61	0.4	1.374	
	Scoring for this item allowed full credit for the correct response and no credit for any other response. In this case we see than only 6% of students were able to correctly round to the nearest hundredth while almost 20% were able to round the number but expressed it in terms of thousandths rather than hundredths. The data do not capture what the majority of students (74%) provided in response to the question but we know that some form of answer was provided. Blanks in every case were given code 9 to differentiate them from incorrect responses.						
	Item:56: 3 /4 + 1 /8 =						A common misconception is when adding fractions with different denominators; students simply add the numerators and the denominators without first finding the equivalent fractions with the same denominator. The same applies to subtraction of fractions with different denominators.
	Common Responses 7/8 Student converts 3/4 to 6/8 and correctly adds the numerators. Equivalent fractions are also correct i.e. 14/16, 21/24, 28/32 ? /8 Student correctly finds the common denominator but not the correct numerator 4 /12 Student adds 3 + 1 to get 4 and 4 + 8 to get 12						Intervention idea: Students might benefit from physical examples of fractions that they can manipulate to demonstrate how fractions are related to one another and to show the results of adding fractions.
Label	**Response**	**Score**	**Count**	**% of tot**	**Pt Bis**	**PV1Avg:1**	*For this item, the average ability (PV1Avg:1) of those whose responses indicated a misconception (8) is a little below but close to those who responded with a partially correct answer (2). It seems to be an identifiable group, suggesting room to focus teaching on this particular error.*
0	Any other response	0	5471	45.41	-0.47	0.021	
1	7/8 (correct)	2	1819	15.1	0.41	1.673	
2	Correct common denominator	1	1562	12.96	0.13	0.896	
8	4/12	0	3196	26.53	0.1	0.663	

	ITEM AND RESPONSE DATA	NOTES AND INTERPRETATION
	Scoring for this item allowed full credit for the correct response, even if it was expressed as an equivalent fraction to the expected 7/8. Partial credit was allowed for responses that showed a correct common denominator was achieved but other errors were made, arriving at an incorrect response. No credit was given for any other response. In this case we note that, while 15% of students are able to answer correctly and another 13% can get to the point of finding the common denominator, one in four students (26%) are doing "straight addition" of the parts of the fraction.	
	Item:78: What time is shown on the clock?  	A common mistake is the interchange of long and short hands
	Common Responses 10 mins to 3 – Student correctly tells this time in numbers or words i.e. ten minutes to three, 10 to 3, 50 mins past 2, etc. 10 mins to, – Student gives an incomplete statement of telling the time correctly in numbers or in words: ten mins to, 10 to. 10.15, ¼ past 10 – Student provides the time as it would be stated if the two hands were reversed, in numbers or in words, quarter past ten, 15 mins past 10.	

Label	Response	Score	Count	% of tot	Pt Bis	PV1Avg:1
0	Any other response	0	7623	63.77	-0.38	0.279
1	10 mins to 3 (correct)	1	3763	31.48	0.36	1.142
2	10 mins to (incomplete statement)	0	352	2.94	0.07	0.947
8	10.15, ¼ past 10 (clock hands reversed)	0	216	1.81	0.03	0.790

For this item, the abilities (PV1Avg:1) of those whose response is incomplete (2) and those whose response indicated the misconception (8) described above are close to those who responded with a fully correct answer but well above those giving a completely incorrect response (0), nevertheless there is some separation among the three groups.

	Scoring for this item allowed full credit for the correct response and no credit for any other response. In this case we see that over 30% of students were able to identify the time correctly and almost all of the rest of the students were not close in their responses, neither inverting the clock hands nor getting a time that was roughly close.

5.3.2 LITERACY

In the two examples following, the item analyses are provided in relation to a small selection of literacy test questions. The Notes and interpretation column explains different aspects of understanding, misunderstanding, errors or lack of understanding that might be indicated by different student responses. Following this analysis, the frequency of common errors or misconceptions are noted and some suitable activities for classroom interventions suggested to address them.

Table 5.2: Summary of item analysis for selected literacy items

<table>
<tr><th colspan="6">ITEM AND RESPONSE DATA</th><th>NOTES AND INTERPRETATION</th></tr>
<tr><td colspan="6">Item:50:
ICE CREAM
I want to eat ice cream
Any colour ice cream
Ice cream in a cone,
Or ice cream in a bowl.

I like ice cream any way.
“Melting on a hot day!” I say,
but it is sweet in my mouth.
And cool in my tummy.

Q: Write the line that tells us what the ice cream tastes like.
NOTE: Has to be the whole line</td><td>In reading a poem, for a question such as, “Write a line used in the poem .....” students are expected to only quote a single line from a stanza.
However, students have responded with a whole sentence which may include two lines instead of a single line or verse. This common “incorrect” (coded 8) answer points to a misconception with elements of poems and language features in general.</td></tr>
<tr><td>Label</td><td>Score</td><td>Count</td><td>% of tot</td><td>Pt Bis</td><td>PV1Avg:1</td><td rowspan="4">Students who do not grasp basic poetic devices have similar ability (PV1Avg:1) as those students who understand the concept.
For this item, responses coded with the label ‘8’ have shown this lack of understanding. The students are a clearly identified group of intermediate ability, on average, between those getting the item correct (1) and those incorrect (0). The much larger group that responded incorrectly (coded with label ‘0’) shows a much lower average ability (PV1Avg:1 value of -0.6), clearly discriminating them from the other two groups, whose responses were coded ‘1’ (correct) or ‘8’ (an identified common error). The PV1Avg:1 values of these two latter groups are relatively close to each other, suggesting similar average abilities.</td></tr>
<tr><td>0</td><td>0</td><td>7229</td><td>60.69</td><td>-0.36</td><td>-0.617</td></tr>
<tr><td>1</td><td>1</td><td>4174</td><td>35.04</td><td>0.33</td><td>0.207</td></tr>
<tr><td>8</td><td>0</td><td>508</td><td>4.26</td><td>0.08</td><td>0.070</td></tr>
<tr><td colspan="7">Suggestions for classroom activities and intervention:

• teach basic elements of poems such as style in simple poems
• help pupils differentiate between a sentence and a line in a poem
• identify lines, stanzas, repeated sounds and words, onomatopoeia (words that imitate a sound);
• make a list of vocabulary in poetry like lines, stanza, rhyme, images and define each term</td></tr>
</table>

ITEM AND RESPONSE DATA						NOTES AND INTERPRETATION
Item:73: Marys book is on the table. Its cover is green.						The student is asked to correctly place an apostrophe in 'Marys'. A common misconception is using an apostrophe in possessive pronouns such as its, hers, theirs in the sentence example, and not in contracted forms such as it's for it is or you're for you are.
Label	**Score**	**Count**	**% of tot**	**Pt Bis**	**PV1Avg:1**	For this item, more than 60% of students placed a frequently used punctuation mark incorrectly ('0', '2' and '8'). Responses coded with labels '2' (apostrophe after 's' in Marys) and '8' (apostrophe is placed in its) are identified to be specific common errors made in the use of such punctuation. The respective PV1Avg:1 values calculated for each label give the average ability of each group. These values suggest that the '8' respondents are closer in average ability to the group that responded correctly than either the '2' or '0' groups.
0	0	3607	55.04	-0.4	-0.286	
1	1	2435	37.15	0.38	0.577	
2	0	50	0.76	0	0.057	
8	0	462	7.05	0.06	0.245	
Suggestions for classroom activities and intervention: • Exposure to the use of apostrophe in a variety of sentences/contexts • Create (together with the students) a Possessive vs Contractions word list • Explore different ways to make singular and plural nouns possessive by adding an apostrophe and an "s" at the end of a word, or just an apostrophe • Practise rewriting groups of words or sentences using possessive nouns • Discuss the reasons for contractions • Have students identify contractions in text and then substitute the two words that the contraction replaces						

5.4 CONCLUSIONS

Coding of the cognitive instrument offers significant potential for additional information that can provide deeper insight into the learning outcomes and learning needs of Pacific Island students. A coding structure, therefore, provides added value to the data and opens further opportunities for PILNA data to be used by education systems, teachers and parents.

This chapter has discussed the technical process of coding, and how item analysis can be applied to understanding reasons for students' responses to certain items. In other words, such analysis leads to a better understanding of student ability and also improves the way items are phrased or presented.

This chapter has also provided a range of examples from both the literacy and numeracy tests for PILNA 2015. These items have been generalised (they are not actual test items) for the purposes of this public report and test security, but they give an indication of the potential for insight into the kinds of learning outcomes and learning needs of the students taking PILNA 2015. The analysis of literacy items also includes how a classroom teacher might address some of the common misconceptions of students about particular literacy skills. The analysis of numeracy items makes similar suggestions based on common misconceptions about mathematical concepts.

CHAPTER 6: PILOT QUESTIONNAIRE DEVELOPMENT FOR COLLECTION OF STUDENT AND SCHOOL DATA

A major purpose of an assessment programme such as PILNA is to derive data and information that can be used to improve the learning outcomes of students in participating countries. Cultural, social and economic factors contribute to student learning contexts. Information collected from questionnaires has the potential to help countries explore connections between how students perform on an assessment and student background, attitudes to schools and approaches to learning. Lack of contextual data was recognised as a gap in data collection in the 2012 PILNA administration. For this reason, a questionnaire was developed by EQAP and piloted in the 2015 PILNA administration. Pilot questionnaires were distributed to a sample of students, teachers and head teachers across the region.

The PILNA questionnaire instruments were piloted with the aim of providing empirical evidence about the relevance, reliability and potential usefulness of the questions selected for inclusion. As a main caveat, it is not possible to develop any conclusions about the responses in the student, head teacher and teacher questionnaires. The primary reason for conducting the pilot process was to test the relevance of questions and then refine them to develop a contextual instrument for full implementation in future PILNA cycles.

The purpose of this chapter is to discuss the pilot process for PILNA 2015. It also includes examples of the kinds of contextual data that can be collected with questionnaires.

6.1 RESEARCH QUESTIONS AND RATIONALE

The PILNA 2015 context questionnaire framework prioritised what information should be collected in a particular assessment administration. It also reflected the purposes of the programme and the capacity to use the assessment data to improve learning outcomes.

Further, decisions about what information to collect were informed by an understanding of how the information can be used to effect changes in policy and practice, and the existence of mechanisms to make those changes.

Specifying research questions makes theoretical assumptions more explicit and clearly indicates what information needs to be collected from pupils, teachers and head teachers. Research questions specific to the three target groups were proposed to guide the development of a context questionnaire framework. The main research question guiding questionnaire development is the following:

1. What factors influence students' achievement in PILNA?

To address this research question, the following sub-questions were included to guide the questionnaire development for students, teachers and head teachers:

a. How do socio-economic factors influence student achievement in PILNA?
b. How do student study habits influence achievement in PILNA?
c. What influence does teacher quality have on student achievement in PILNA?
d. How do school management and leadership influence student achievement in PILNA?

6.2 DEVELOPMENT OF PILOT QUESTIONNAIRES

Pilot questionnaires for PILNA 2015 were developed by EQAP in consultation with countries participating in PILNA. The draft pilot questionnaires were reviewed and edited by research staff, which included alterations to the structure and coding for some of the questions.

The questionnaire component of PILNA collects background and contextual data from various sources. The information can be a powerful tool in providing explanations for the outcomes of the assessment's cognitive component. This information enables a more in-depth understanding of the observed test outcomes (student learning outcomes), and the implications for designing interventions to improve student learning outcomes.

A range of background factors are associated with different levels of student performance, and some factors may be amenable to change as a result of policy decisions and practical arrangements of schooling. Background information for PILNA 2015 was gathered from different levels:

1. Student level information includes information about gender, age, schooling history, interests and motivations, and family and home environment.
2. School level information includes information about teaching and learning practices, qualifications, expertise and professional development practices of teaching staff, school culture, resourcing and organisational practices of the school, and school-level policies.

System level information includes broader contextual factors about the operation of the school within the education system, including funding, staffing levels and accountability arrangements, quality assurance arrangements, and system support.[13]

Throughout the process of questionnaire development, extensive reviews and discussions were conducted with Pacific education and assessment experts. The development of the pilot questionnaires involved three processes.

1. Development of first-draft material and review by PILNA national coordinators and experts.
2. Review by experts and field trial conducted in the 13 participating PILNA countries.
3. Analysis of the field trial results, followed by a final selection and refinement of the main survey items. The aim of this phase was to finalise the content of the student, teacher and head teacher questionnaires.

Throughout each review of the pilot questionnaires, the experts used the following criteria to select proposed item material:

a. relevance with regard to the PILNA assessment framework;
b. appropriateness for the national contexts of the participating countries;
c. psychometric properties of items designed to measure latent traits postulated in the initial formulation and found in the field trial data.

The analysis of the field trial data provided empirical evidence on the quality of the item material and informed the selection of the main survey material.

6.3 THE PILNA PILOT QUESTIONNAIRES

The PILNA background questionnaires were administered to students, teachers and head teachers at sampled and census schools as a pilot.

1. The student questionnaire generated information about students and student backgrounds in order to gather a better understanding of student learning outcomes on the cognitive assessment.
2. The teacher questionnaire generated information about the teaching and learning environment of sampled students, and how that environment might contribute to understanding student learning outcomes on the assessment.
3. The head-teacher questionnaire generated information at a higher level of school and (potentially) system organisation that may help to explain observed differences in assessment outcomes across the sampled students.

6.4 PILOT SAMPLE AND ADMINISTRATION

Pilot questionnaires for students, teachers and head teachers were trialled in each country as part of the PILNA 2015 implementation. The definition of the target population and the sampling strategy for selecting students was framed around the following method:

All Year 4 and Year 6 students at selected schools: A (random) sub-sample of students at three schools (alternatively at the whole school) were invited to complete the questionnaire. Schools were selected from three different locations as relevant per country: urban, rural and remote. Two countries participating in PILNA had three or less schools nationwide, and, therefore, a census of students was invited to complete the questionnaire. All Year 4 and Year 6 students who participated in PILNA from the selected schools were invited to complete the questionnaire.[14]

13 Some system level information is gathered from head teacher questionnaires, but related to local systemic support such as accountability and quality assurance. These questions, however, are not included in the examples in the latter half of this chapter.

14 There should be at least 100 responses per item to test the psychometric properties of each item.

Definition of the target population and the sampling strategy for selecting teachers and head teachers was framed around the following method:

All teachers and head teachers at the target grade or school: All or a (random) sub-sample of teachers at the target grade (alternatively at the whole school) was invited to complete the questionnaire. This approach is useful when the primary aim is to gather information about the whole school context (in addition to what is obtained from head teachers), and includes larger teacher sub-samples within selected schools. The link between student, head-teacher and teacher data is at the school level.

For pilot studies it is recommended not to limit the teacher survey to class teachers of sampled students in order ensure sufficient numbers of teachers in the pilot sample (e.g. by including all target grade teachers at piloted schools).

The pilot questionnaires were administered at the same time as the PILNA literacy and numeracy assessments. The completed questionnaires were returned to EQAP where they were scanned into a database.

6.5 ANALYSIS OF PILOT QUESTIONNAIRES

As mentioned at the outset of this chapter, the primary reason for analysing the data from the pilot questionnaires was to test the relevance of questions to the PILNA assessment programme, and the psychometric properties of the questionnaire items. In this case, the purpose of the pilot questionnaire was to test the content and structure of questions that can be included in future cycles of PILNA.

The questionnaire was a pen and paper instrument, and was distributed to students, teachers and head teachers during the PILNA 2015 administration. The sampling and census methods described above were applied to the questionnaire administration.

The results included in this chapter are only frequencies and distributions. As the questionnaires are in pilot form, we have not presented any results where contextual factors are correlated to achievement data from the cognitive instrument. However, correlational analyses will be used to aid question selection for future PILNA cycles. The main purposes of presenting such results are to provide insight into the characteristics of the sample population, and to demonstrate the potential that contextual information can provide on the nature of educational systems.

6.6 LIMITATIONS IN PRESENTING RESULTS

There are limitations in presenting the results of pilot questionnaire data, particularly as there were large differences in responses across the countries of the region, both in terms of the number and characteristics of the respondents. These differences might be representative of sample design for this pilot (particularly for student questionnaires), or perhaps it suggests that there are significant differences in the educational experiences across the Pacific Island countries. While only a small proportion of schools were surveyed in some countries, in two countries the pilot was a census, whereby all students, teachers and head teachers were invited to participate in the survey. For this reason, any findings presented may be biased towards a particular country context, and may not be representative of the region as a whole.

All head teachers at PILNA schools were asked to participate in the survey, and there was a high percentage of usable data. At the same time, while there was a large amount of usable data from students, only students in a convenience sample of up to three schools per country were invited to participate in the survey. This means there is extremely limited scope in the ability to provide a meaningful discussion on student related responses, and only basic characteristics of students are included in this chapter. Limited conclusions can be made about interpreting the nature of the student, teacher and head teacher responses.

6.7 STUDENT QUESTIONNAIRES

The student questionnaire collected data about student characteristics, family and home background. It asked students what language they speak at home, parent/guardian employment and home support for studying. It also collected information about student homework habits and their perceptions of school.

The field trial student questionnaire included a total of 28 questions and was administered to a sample of 1,800 Year 4 and Year 6 students.[15] The characteristics of these students are presented in Table 6.1.

Data were collected from 13 countries across the region. The gender distribution of the respondents was roughly equal, with slightly more girls (51%) completing the questionnaire. The majority of respondents were aged between 9 and 13 years (78%), with smaller proportions aged 8 years and younger (8%) and 13 years and older (14%). This age range is consistent with the year levels sampled, with Year 4 pupils comprising 52% of the data, and Year 6 pupils comprising 48% of the data.

15 Some PILNA countries had total student population sizes in Year 4 and Year 6 that were less than 100, and in those cases, a census (all students) were invited to complete pilot student questionnaires.

Table 6.1: Student characteristics from the PILNA 2015 pilot questionnaire

Student respondent characteristics		Number of students	Percentage of students
Gender	Female	560	51
	Male	534	49
Age	8 and younger	82	8
	9	211	19
	10	208	19
	11	242	22
	12	194	18
	13 and older	146	14
Class year	Year 4	576	52
	Year 6	530	48

The amount of missing data for each question varied. Percentages represent the proportion of non-missing responses.

6.8 HEAD TEACHER QUESTIONNAIRES

The head teacher questionnaire collected data about school management and leadership. Head teachers were asked about the qualifications and experience of teachers in their schools, and about how long they had been a head teacher. They were also asked to comment on professional development and the management of school resources.

Head teacher questionnaires were distributed to all schools in the 13 countries participating in PILNA. Of those who completed a questionnaire, usable data were collected from 405 respondents. The characteristics of these head teachers are presented in Table 6.2.

Overall, slightly more female (53%) than male head teachers responded to the questionnaire. The age profile shows that approximately half of all respondents were over the age of 45, with only a small proportion (4%) 30 years or younger in age. In general, approximately half of all respondents were relatively inexperienced, having been a head teacher for five years or less. The most common highest academic qualification reported was a diploma (40%), followed by a tertiary certificate (23%). The vast majority of all head teachers reported completing a teacher training programme (93%).

Table 6.2: Head teacher characteristics from the PILNA 2015 pilot questionnaire

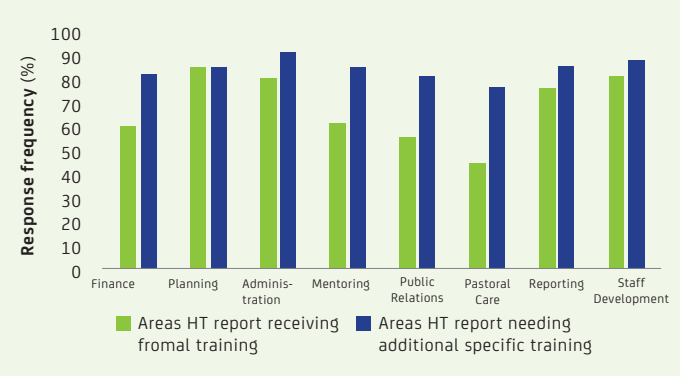
HT respondent characteristics		Number of head teachers	Percentage of head teachers
Gender	Female	211	53
	Male	190	47
Age	25 and younger	4	1
	26-30	12	3
	31-35	28	7
	36-40	69	17
	41-45	94	24
	46-50	90	23
	Older than 50	103	26
Number of years as a head teacher	5 and below	196	51
	6-10	101	26
	11-15	41	11
	16-20	27	7
	Greater than 20	23	6
Highest academic qualification	High School Certificate	49	13
	Tertiary Certificate	88	23
	Diploma	154	40
	Bachelor	73	19
	Higher Degree	19	5
Completed a teacher training programme	Yes	360	93
	No	28	7

The amount of missing data for each question varied. Percentages represent the proportion of non-missing responses.

Head-teachers were asked if they had completed training in leadership and management. Of those who answered this question, just over three-quarters indicated that they had received training in leadership and management. These teachers were asked in which areas they had received training. Figure 6.1 indicates that higher proportions indicated they received training in planning (85%), staff development (81%), administration (80%) and reporting (76%). Lower proportions reported receiving training in mentoring (61%), finance (60%), public relations (55%) and pastoral care (44%).

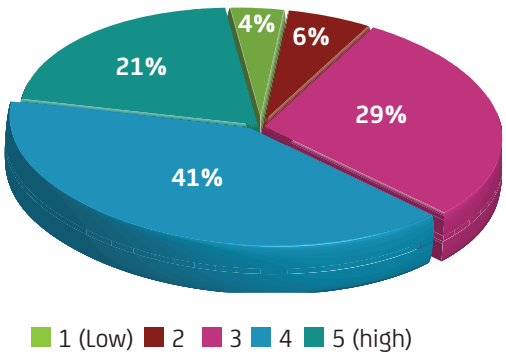
All head teachers (including those who indicated that they had not received formal training in leadership and management) were asked the areas they felt they needed specific additional training in. Figure 6.1 shows that a large proportion (more than 75% of all respondents) indicated they needed additional training in all eight areas. Administration (91%) and staff development (87%) were the areas that attracted the highest proportion of respondents.

Figure 6.1: Head teacher training in management and leadership, pilot questionnaire for PILNA 2015



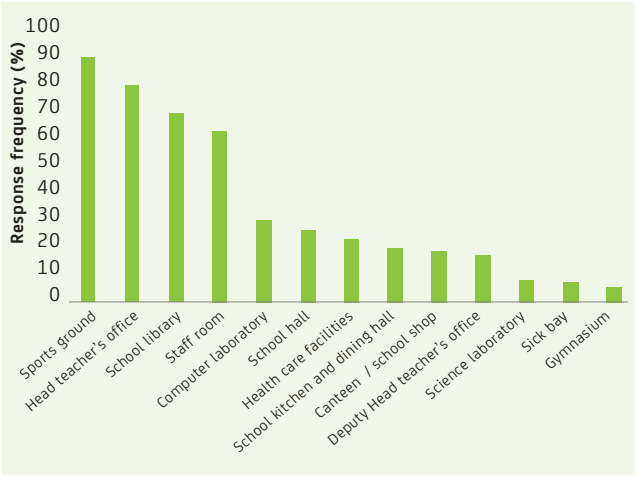
Head teachers were asked to rate their level of satisfaction in their role at their school, using a scale ranging from 1 to 5, where 1 indicates that they are not at all satisfied, and 5 indicates that they are very satisfied. Figure 6.2 shows that approximately 10% of respondents selected 1 or 2, indicating a low level of satisfaction in their jobs; 21% indicated that they were very satisfied (selecting 5), and 41% indicated fairly high levels of satisfaction (selecting 4).

Figure 6.2: Level of satisfaction in head teacher role, pilot questionnaire for PILNA 2015



Head teachers were asked to indicate the facilities that were present at their school from a list of 13 items: a head teacher's office, deputy head teacher's office, staff room, school library, science laboratories, computer laboratory, school kitchen and dining hall, school hall, gymnasium, sports grounds, canteen / school shop, health care facilities and a sick bay. Figure 6.3 indicates that the four facilities that were most commonly reported at schools were a sports ground (87% of all schools), a head teacher's office (78%), a school library (67%) and a staff room (61%). All other facilities were present in less than 30% of schools. A particularly low proportion of schools had a science laboratory (8%), a sick bay (7%) and a gymnasium (5%).

Figure 6.3: Reported school facilities, pilot questionnaire for PILNA 2015



Respondents to the questionnaire were asked a series of questions relating to teacher management. Table 6.3 shows that just under a half of all head teachers (49%) indicated that they evaluate the performance of their teachers at least once a term, with smaller proportions indicating they did this two to three times a year (26%) and once a year (10%). Only a very small proportion of respondents indicated this was done less than once a year (1%), or that there was no evaluation (2%). Approximately one in eight respondents suggested that evaluation was done as required (12%).

With regards to monitoring of teacher attendance, head-teachers were nearly universal in their responses that some sort of monitoring takes place, most commonly an attendance register that teachers need to sign (90%). A smaller proportion of schools require teachers report to the head teacher/deputy head teacher (6%), or some other system is in place (3%).

Professional development of teachers appears to be a high priority in schools in the Pacific Island region, with professional development plans for teachers reportedly existing in 85% of all schools.

Table 6.3: Head teacher management of teachers, pilot questionnaire for PILNA 2015

Management of teachers		Number of head teachers	Percentage of head teachers
How often do head teachers evaluate the performance of teachers in their school?	At least once a term	192	49
	Two or three times a year	102	26
	Once a year	38	10
	Less than once a year	2	1
	As required	47	12
	No evaluation	9	2
How is teacher attendance monitored?	Teachers sign an attendance register	353	90
	Teachers report to head teacher/ deputy head teachers	25	6
	Other	13	3
	Teacher attendance not recorded	2	1
Is there a professional development plan for teachers?	Yes	328	90
	No	58	15

The amount of missing data for each question varied. Percentages represent the proportion of non-missing responses.

6.9 TEACHER QUESTIONNAIRES

The teacher questionnaire collected data about teachers' age, experience and access to resources. In particular, the questionnaire asked teachers about their qualification and length of service, as well as opportunities for and engagement with professional development opportunities. It also asked teachers about challenges in particular aspects of their teaching and assessment practices, their access to teaching resources, and their interaction with their school communities.

Teacher questionnaires were targeted towards teachers of years 4 and 6 and were received from all participating PILNA countries. The characteristics of those that responded are presented in Table 6.4.

The age profile of respondents appears to be spread out, with relatively large proportions of teachers aged between 21 and 30 (23%), between 31 and 40 (37%), between 41 and 50 (26%) and above 50 (14%). However, overall, teachers were relatively inexperienced, with 98% of respondents having 10 years of experience or less.

A spread of highest academic qualification was also observed. While 50% had a diploma as their highest education, 20% had a high school certificate, 20% had a tertiary certificate, while 10% had a bachelor's degree. Only a small proportion (1%) had a higher degree. A large proportion of teachers had completed a teacher training programme (84%).

Table 6.4: Teacher characteristics, pilot questionnaire for PILNA 2015

Teacher respondent characteristics		n	%
Age in years	20 or below	8	1
	21–25	55	8
	26–30	97	15
	31–35	138	21
	36–40	106	16
	41–45	115	18
	46–50	50	8
	50 or more	90	14
Years of experience	5 or below	271	41
	6-10	373	57
	11-15	8	1
	21-25	2	0
	25 or more	3	1
Highest academic qualification	High school certificate	124	20
	Tertiary certificate	125	20
	Diploma	312	50
	Bachelor	61	10
	Higher degree	4	1
Completed a teacher training programme	Yes	532	84
	No	99	16

The amount of missing data for each question varied. Percentages represent the proportion of non-missing responses.

Teachers were asked to indicate aspects of both literacy and numeracy that they find difficult to teach. Figure 6.4 indicates that teachers that responded to the survey have particular difficulty with teaching composition writing, as this was the most common aspect of literacy selected (63% of all respondents). Other literacy aspects that teachers reported finding difficult to teach were reading – comprehension; reading – grammar and punctuation (both 50%); oral language – speaking (48%) and listening (46%).

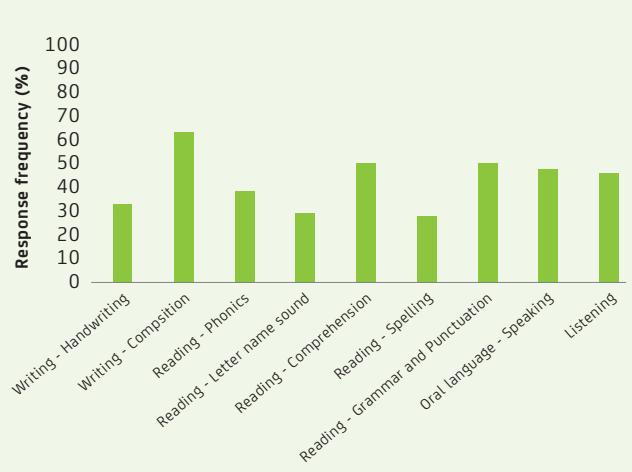


Figure 6.4: Difficult aspects of teaching literacy, pilot questionnaire for PILNA 2015

Figure 6.5 shows the aspects of teaching numeracy that respondents indicated were difficult to teach. For numeracy, problem-solving was the most common aspect that teachers rated as difficult to teach (61%). Geometry (45%) and patterns and algebra (40%) also attracted relatively high proportions of responses.

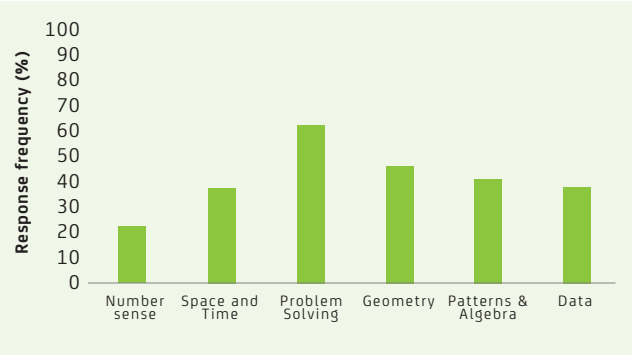


Figure 6.5: Difficult aspects of teaching numeracy, pilot questionnaire for PILNA 2015

Teachers were asked to indicate their level of confidence in using their pre-service training in assessment in their work on a scale from one to five, where one suggests that they are not confident, and five suggests that they are extremely confident. Of those teachers who responded to the survey, Figure 6.6 shows that only 4% of respondents indicated that they were not confident (selecting a score of one) in using their pre-service training in assessment, whereas the majority (64%) indicated that they were relatively confident (selecting a score of four or five on the scale).

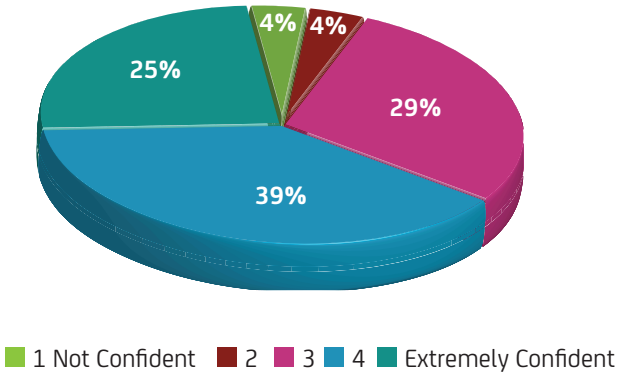


Figure 6.6: Confidence in using pre-service training in assessment, pilot questionnaire for PILNA 2015

Teachers were asked if they had participated in professional development activities in the past three years. Figure 6.7 shows that approximately two out of three teachers (67%) indicated that this was the case.

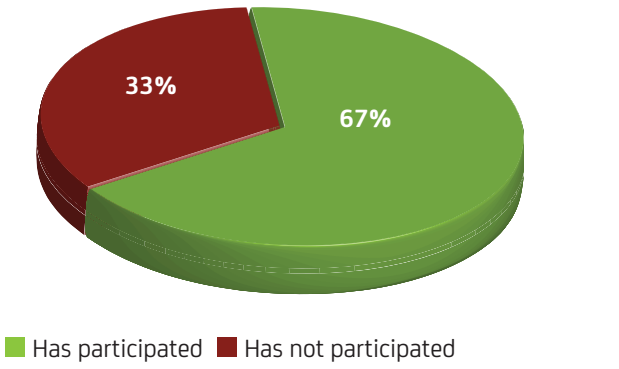


Figure 6.7. Participation in professional development activities, pilot questionnaire for PILNA 2015

6.10 CONCLUSION: USE OF THE PILOT QUESTIONNAIRE RESULTS FOR PILNA

The three questionnaire instruments described here provide a long-term frame of reference. They have generated a set of concepts, scales and indices against which student learning outcomes can be analysed and understood in future cycles of PILNA. The student, teacher and head teacher pilot questionnaires were developed in a multi-stage review process that included a field trial in all 13 of the participating PILNA countries with the input of national and regional education experts.

As noted earlier in this report, the results from student performance in literacy and numeracy are raw results and differences, unadjusted for any relevant background factors that might explain the reported observations. For example, we might find that the differences observed by school type (government and non-government schools) might vanish if we adjusted for some background variables, such as home support for studying, language spoken at home, parent/guardian education, etc. In the future, information collected from questionnaires could usefully contribute to determining the extent the observed differences in student learning outcomes can be attributed to the variable, such as school type or gender.

Critically, the purpose of the pilot questionnaire is to provide a stable and consistent framework within which performance changes over time can be analysed, differences within the sampled populations can be monitored and understood, intervention strategies can be designed and evaluated, and learning improvements can be forged.

CHAPTER 7 - CONCLUSIONS, SUMMARY AND RECOMMENDATIONS

The 2015 PILNA regional report provides an in-depth analysis of the numeracy and literacy assessment outcomes of Year 4 and Year 6 students in 13 Pacific Island countries. It also reports on trends in student achievement in literacy and numeracy between 2012 and 2015.

The data in the regional report included information on two major subgroups – gender and school of locality (urban, rural, remote or very remote) is not included in the regional authority – in order to get comparative information about learning outcomes. The categorisation and small islands states reports because of the differing definitions of locality in each of the PILNA countries.

The data suggest that there is significant improvement across the region in numeracy at both Year 4 and Year 6, as well as improvement in numeracy between 2012 and 2015. On the other hand, performance in literacy is varied, with Year 4 students showing improvement in literacy, but Year 6 students showing minimal change in literacy performance.

Looking at the subgroups, girls outperformed boys in both numeracy and literacy at both year levels, as well as between 2012 and 2015. In the school authority subgroup, student performance was varied. In numeracy, Year 4 students in non-government schools outperformed their counterparts in government schools, but Year 6 students in government schools outperformed their counterparts in non-government schools. In literacy, students in both year levels in non-government schools outperformed students in government schools, but the variation was minimal.

This report also describes two innovations that will contribute to the improvement of a long-term assessment programme in the region. One innovation is the implementation of a system of coding rather than scoring, and other is the inclusion of context questionnaires for students, teachers and head teachers. Both coding and pilot questionnaires were trialled in 2015 and evidence from the field trials will contribute to design framework for future PILNA cycles. In addition a coding scheme and context questionnaires.

7.1 METHODOLOGICAL FRAMEWORK

The overall methodology of PILNA 2015 provides a comparative analysis of data with the Pacific regional benchmarks, student performance on PILNA 2012, and student performance of countries in the region as a whole. An important element is that country-to-country comparison is NOT a component of the programme, as explicitly directed by FEdMM in 2014. Chapter 2 discussed key methodological inputs for PILNA 2015.

As described at the outset of this report, the Pacific is one of the world's largest and most diverse regions. Given the extreme variations across the countries participating in PILNA, the sampling design is a complex process.[17] It uses a census approach for the relatively smaller countries, and a sampling approach for a number of the larger countries included in the study.[18] One larger country, population-wise, requested that a census approach be applied to its PILNA administration.

Participating countries were given the opportunity to have the 2015 PILNA instruments translated, in line with the definition of literacy in the regional benchmarks. After considering their individual language policies and the language of instruction/testing at both Year 4 and Year 6, nine countries opted for a translation.

Student outcomes were reported on a single uniform metric scale that was constructed to achieve two main goals: first, to provide descriptions of what students can do at various points along the scale; and second, to show student achievement by year level in a way that can be reported and interpreted consistently across all participating populations. A set of new proficiency scale levels was also developed, based on the item-to-skill mapping and placing the items on a Guttmann structure (i.e. ordering the items by difficulty and establishing level cut-offs based on skill and content grouping of the items). The proficiency scale levels give education stakeholders information about what students know and can do at particular points in their learning.

17 The sampling process is documented in detail in the 2015 PILNA Technical Report.
18 The smaller countries include Cook Islands, Niue, Palau, Tokelau and Tuvalu. The larger countries include Federated States of Micronesia, Papua New Guinea, Republic of the Marshall Islands, Samoa, Solomon Islands, Tonga and Vanuatu. Kiribati requested a census approach.

Finally, expected levels in literacy and numeracy were developed to provide a reference point for the countries to indicate the minimum standard of achievement for students who have gone through four and six years of schooling. The process of setting the expected levels entailed discussing the learning outcomes on the proficiency scale, focusing on the specific skills and knowledge that are represented at each level of the scale. The expected levels were then finalised, based on how the learning outcomes mapped the regional benchmark indicators in literacy and numeracy.

7.2 REVIEW OF THE KEY FINDINGS

This report has discussed in depth the literacy and numeracy performance of students who have completed four and six years of schooling between 2012 and 2015. In general results showed improvement in Year 4 and Year 6 in numeracy, and in Year 4 literacy, but not in Year 6 literacy. This section reviews the overall findings of PILNA 2015.

Notable improvement in numeracy achievement across the region

In total, the results of 11,289 Year 4 students and 14,292 Year 6 students were analysed for the 2015 PILNA numeracy test. Student performance in numeracy achievement improved, both in student progression between year levels and in the proportion of students performing at higher proficiency levels. Students at both year levels also demonstrated improved performance in all strands of the numeracy domain – numbers, operations, and measurement and data – between 2012 and 2015.

Greater proportions of students at both Year 4 and Year 6 are reaching the higher proficiency levels and are at or above the expected proficiency levels. And, as would be expected, substantial improvement is observed as students progress from Year 4 to Year 6.

In addition to improved performance in the general domain of numeracy, student mean performance also improved in all strands of the domain. Year 4 students showed the highest proficiency in measurement and data, while Year 6 students showed the highest proficiency in numbers.

Similar numbers of boys and girls in Year 4 and Year 6 participated in PILNA 2015. At the regional level, girls outperformed boys in numeracy in both Year 4 and Year 6, although the difference in performance between boys and girls was minimal. Girls also performed better than boys at both year levels in all the strands. There was a slight difference in the distribution of boys and girls across the proficiency levels: a slightly higher proportion of girls than boys in the upper proficiency levels in both Year 4 and Year 6, and there were more boys than girls in the lower proficiency levels.

Finally, the comparative mean performance in numeracy of students in government and non-government schools was inconsistent. In Year 4, students from non-government schools slightly outperformed students from government schools in the overall numeracy domain, as well as in all the strands, based on the average regional achievement in numeracy. However, Year 6 students from government schools slightly outperformed students from non-government schools.

Some improvement in literacy in the distribution of students across the proficiency levels from 2012 to 2015

In total, the results of 12,194 Year 4 students and 15,807 Year 6 students were analysed for the 2015 PILNA literacy test. Improvement in literacy across the region has been inconsistent. There was a small improvement in student performance from Year 4 to Year 6 with more students performing at higher proficiency levels and fewer students performing at lower proficiency levels. The average mean performance in literacy of students in Year 4 improved between 2012 and 2015, while there was no change for students in Year 6 between 2012 and 2015. Students in both year levels performed best in the strand of reading, followed by language features and writing.

However, the improvement across proficiency levels is not consistent between 2012 and 2015. On the one hand, there were fewer students in the lowest proficiency levels and more students in the middle range (levels 3–6) of the proficiency scale in years 4 and 6. At the same time, there was a decline in the number of students achieving at the highest proficiency levels between 2012 and 2015. This signifies somewhat uneven progression in literacy between the two year levels.

Increases in achievement were evident in all the strands (reading, language features and writing), as well as in the general domain of literacy. The Year 4 average mean performance in literacy increased between the 2012 and 2015 cycles of PILNA. Statistically, however, the overall performance in Year 6 revealed flat, or minimal, change between 2012 and 2015.

In terms of gender, findings showed a notable difference in the distribution of boys and girls across the proficiency levels in literacy. There were more girls than boys in the higher proficiency levels, and more boys than girls in the lower proficiency levels. In each proficiency level at or below the expected level for Year 4 (level 4) and Year 6 (level 5), there were proportionally more boys than girls. At the same time, there were proportionally more girls than boys among those who achieved above the expected proficiency level in literacy. Generally, students from non-government schools slightly outperformed students from government schools, although the difference was relatively small. In both Year 4 and Year 6, there was a slightly higher proportion of non-government school students in the upper proficiency levels compared to students from government schools. Across all proficiency levels, students in non-government schools showed a higher level of achievement in literacy than students in government schools.

Innovations in PILNA 2015 have the potential to enrich data on student learning outcomes in literacy and numeracy

In 2015, two innovations were piloted - a coding scheme and context questionnaires - that enable a deeper understanding of student learning outcomes. A coding scheme provides teachers, head teachers and other education stakeholders with additional information about why students might respond to particular items in an assessment. By learning more about student misconceptions in responding to items, a teacher may be able to develop interventions or classroom tasks that may help students improve learning outcomes.

Context questionnaires enable more in-depth investigation about background factors of students, teachers and head teachers that may or may not impact student learning outcomes. As noted throughout this report, information collected from context questionnaires might contribute to better understanding of the observed differences from cognitive instruments.

Finally, the development of a regional uniform metric has enabled trends in literacy and numeracy learning outcomes to be monitored over time. In addition, a proficiency scale for literacy and numeracy has been developed for PILNA, which allows education stakeholders to better understand the skills that students develop at particular points in their learning progression. The findings from the cognitive instruments (literacy and numeracy assessments) and the development of coding, questionnaires and a regional uniform metric create a framework for recommendations supporting the future of a PILNA program.

7.3 RECOMMENDATIONS

Tracking the evolution of student learning outcomes in literacy and numeracy over time can help the Pacific region monitor how students are improving in relation to the regional benchmarks approved by the FEdMM in 2007. Importantly, PILNA addresses targets identified in SDG 4 by analysing and measuring results that provide evidence of education quality for national governments and regional organisations to develop interventions that have the potential to support students in improving their skills in literacy and numeracy. The following recommendations are broad and applicable across the region. Action on any of these recommendations could be taken up by individual countries; or, perhaps two or more countries could work in partnership to develop interventions or frameworks to work toward improving student learning outcomes.

Educational stakeholders are advised to review PILNA evidence and trends between 2012 and 2015 both regionally and nationally, and consider intervention strategies for students performing at the lower end of the proficiency scale, particularly in literacy. Data provided at PILNA country and regional levels provides a robust evidence base to support decision-making and policy development at the system, school and, potentially, classroom levels.

To make certain that results reach the classroom level for targeted intervention, education authorities are advised to expand their dissemination approaches when reporting the results of the study, making certain that results reach the classroom for targeted intervention. At the country level, PILNA has also investigated the performance of students based on school location. Again, PILNA provides a key source of data on student learning outcomes that could support potential intervention strategies.

Education stakeholders are strongly encouraged to identify intervention strategies that improve the achievement of boys, especially in literacy. In the process of identifying strategies, it is recommended that deeper analysis of PILNA regional and national results or other data be undertaken in an effort to understand the underlying issues facing boys in literacy in the region. Targeted intervention should be designed based on evidence from a range of sources, with PILNA providing a key source of data on student learning outcomes.

Education stakeholders and EQAP are strongly encouraged to adopt the implementation of context questionnaires as part of a long term assessment programme. The current PILNA results on student performance in literacy and numeracy are raw results and differences, unadjusted for any relevant background factors (such as home support for studying, language spoken at home, parent/guardian education, etc.) that might explain the reported observations. Questionnaires enable deeper investigation of difference observed by gender, school type or school location. In the future, information collected from questionnaires could usefully contribute to

determining the extent the observed differences in student learning outcomes can be attributed to the variable, such as school type or gender.

Education stakeholders and EQAP are strongly encouraged to adopt the implementation of a full coding scheme. A coding process provides information about why some incorrect responses are more frequently provided by students than others. This process has the potential to enable teachers and head teachers to understand how and why their students may be responding to questions in particular ways. Such information can be shared with classroom teachers who can use the data to address misconceptions by students on specific topics.

Regional and national education leaders and FEdMM are strongly encouraged to adopt the use of a regional uniform metric as a way to track progress and trends in student learning outcomes. Measuring learning outcomes on a proficiency scale enables all education stakeholders – teachers, students, parents, local, national and regional authorities – to gather evidence about what students know and can do at a particular stage in their learning development.

Regional education stakeholders are strongly encouraged to support an ongoing PILNA that has the power to provide much more robust evidence to policymakers with richer data from which to develop policies and intervention strategies to improve student learning outcomes. Innovations such as coding, context questionnaires and the developing of a regional uniform metric enable policymakers to explore in-depth the data about student learning outcomes and make decisions about aspects of a country's education situation.

Education stakeholders are advised to investigate ways in which the robust and valid data provided by PILNA can support the improvement of student learning outcomes. Government commitment can provide support and guidance to teachers in translating data into useful information for better results in students' achievement. A variety of reports pitched at different stakeholders (parents, teachers, students, provincial authorities and national authorities) has the potential to provide broad community and political support. This recommendation also has the potential to provide more in-depth information about student learning outcomes and student background in the future.

This 2015 report has provided an analysis of the literacy and numeracy skills of students who have completed four and six years of formal schooling. PILNA developed a regional uniform metric, and thereby explored changes in student achievement in the Pacific over time, between 2012 and 2015. The analysis of trends over this second cycle of PILNA has the potential to enable policy-makers to make informed, evidence-based decisions about how to improve the learning outcomes of students across the Pacific region.

REFERENCES

Altinok, N. (2012). A new international database on the distribution of student achievement. Background paper for EFA Global Monitoring Report 2012.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov,P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. and Japel, C. (2007). School readiness and later achievement. Developmental Psychology 43(6), pp. 1428–1446.

Guttman, L. A. (1950). The basis for scalogram analysis. In Stouffer, S.A., Guttman, L.A. and Schuman, E.A., Measurement and prediction. Volume 4 of Studies of social psychology in World War II. Princeton: Princeton University Press.

Lewin, K. M. (2007). Improving access, equity and transitions in education: Creating a research agenda. Brighton, UK, Consortium for Research on Educational Access, Transitions and Equity, University of Sussex. (CREATE Pathways to Access Research Monograph, 1).

OECD (2014). PISA 2012 Results: What students know and can do – Student performance in mathematics, English and science (Volume 1, Revised). Paris: OECD.

Richmond, M., Robinson, C. and Sachs-Israel, M. (Eds) (2008). The global literacy challenge: A profile of youth and adult literacy at the mid-point of the United Nations Literacy Decade 2003–2012. Paris: UNESCO).

South Pacific Board for Educational Assessment (2012). Pacific Islands Literacy and Numeracy Assessment (PILNA) 2012. Regional report. Suva, Fiji: SPBEA.

UNDP (2015). Sustainable Development Goals. Available at https://sustainabledevelopment.un.org/.

UNESCO (2015). Pacific education for all: 2015 Review. Apia: UNESCO.

UNESCO (2012). EFA global monitoring report. Youth and skills: Putting education to work. Paris: UNESCO.

UNESCO (2006). EFA global monitoring report. Literacy for life. Paris: UNESCO.

UNESCO (2005). Aspects of literacy assessment: Topics and issues from the UNESCO Expert Meeting. Paris, June 10–12. Paris: UNESCO.

APPENDIX

1.1 THE PACIFIC REGIONAL BENCHMARKS FOR LITERACY AND NUMERACY

PACIFIC DEFINITION OF LITERACY:

"Knowledge and skills necessary to empower a person to communicate through any form of language of their society, with respect to everyday life"

A person is considered to be functionally literate if she/he has acquired the necessary knowledge and skills to be able to communicate effectively through any form of language of their society, with respect to everyday life.

The literacy status of a person between the ages of 6 to 14 years will be determined nationally and regionally (if required) by referencing his/her literacy skills to the indicator outlined below. However, a person is considered to be functionally literate if he/she has completed four years of formal education and has met the indicator outlined for Year 4.

SKILL COMPONENT	YEAR 4	YEAR 6
	INDICATORS (9 – 10 YRS)	INDICATORS (11 – 12 YRS)
LISTENING	Listen, understand, respond and question texts	Listen, understand, respond and critically question texts/genre
SPEAKING	Understand, speak, question and respond in various genres/texts	Understand, speak, critically question and respond in various genres/texts
WRITING	Write and present ideas in a variety of genres/texts that demonstrate the basic use of writing mechanics	Writing and presenting ideas including critical analysis in a variety of genres/texts that uses writing mechanics (appropriately).
READING	Read, understand, question and respond to a variety of genres/texts	Read, understand, question and respond critically to a variety of genres/texts

PACIFIC DEFINITION OF NUMERACY:

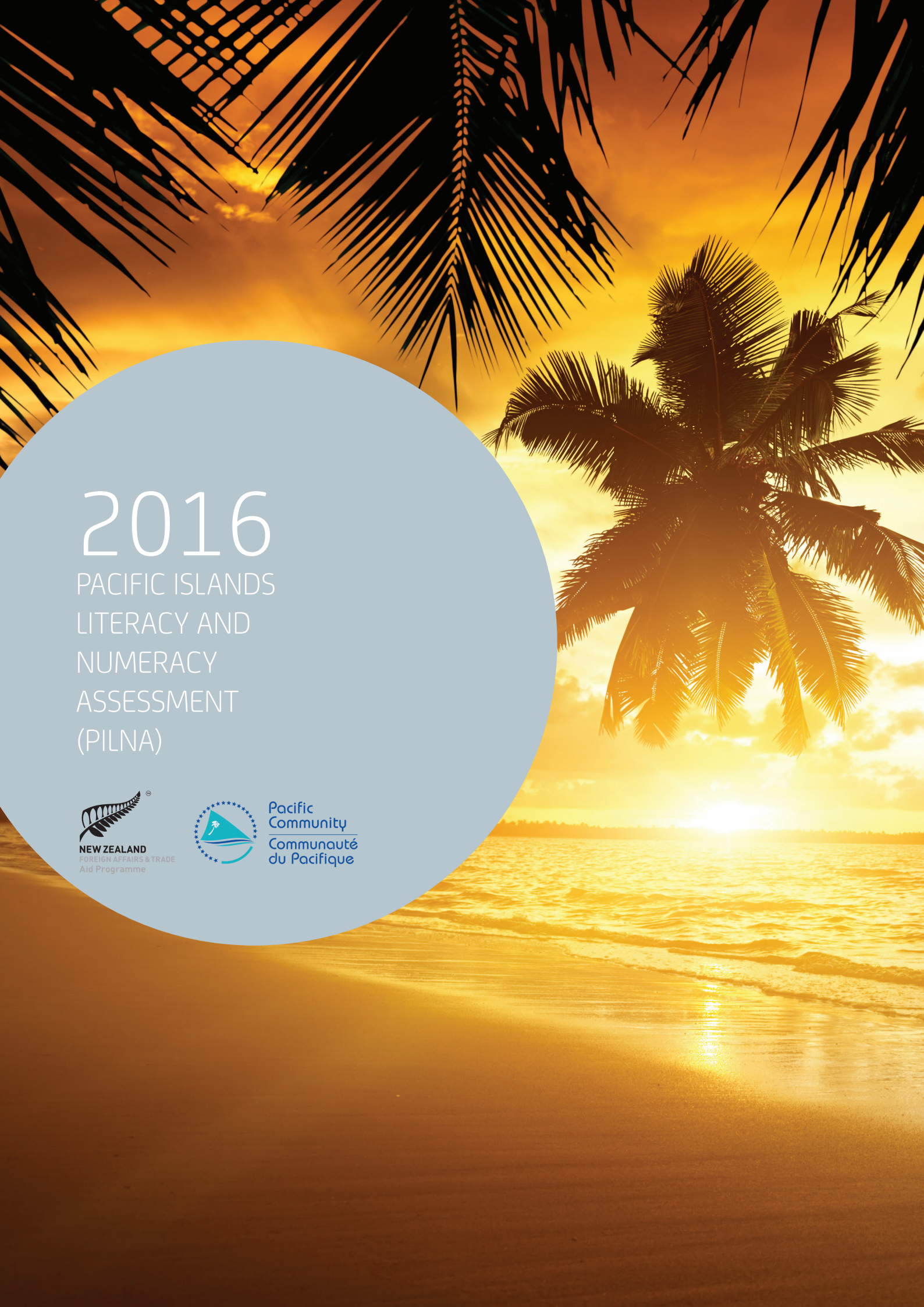
"Knowledge and skills necessary to empower a person to be able to use numbers in mathematical processes, as well as the language of mathematics, for a variety of purposes, with respect to everyday life"

A person considered to be functionally numerate is therefore someone who has acquired the necessary knowledge and skills to be able to use numbers effectively in mathematical processes, as well as the language of mathematics, for a variety of purposes in everyday life not only within the society he/she lives but beyond.

The numeracy status of a person between the ages of 6 to 14 years will be determined nationally and regionally (if necessary) by referencing his/her numeracy skills to the benchmarks outlined below. However, a person is considered to be functionally numerate if he/she has completed four years of formal education and has met the numeracy benchmark outlined for Year 4.

SKILL COMPONENT	ELEMENTS	YEAR 4 INDICATORS (9 – 10 YRS)	YEAR 6 INDICATORS (11 – 12 YRS)
NUMBERS	Counting Objects	Represent numbers using numerals from 1 to 999	
	Recognising Numbers	Identify and write numbers in numerals and in words up to 999;	Identify and write numbers in numerals and in words up to 9999;
	Place value	Identify and write 3-digit whole numbers and decimal numbers up to 2 decimal places;	Identify and write 4-digit whole numbers and decimal numbers up to 3 decimal places; Round off numbers up to 2 significant figures and 2 decimal places;
	Fractions & Percentages	Relate parts of an object to the whole; Identify denominator & numerator in a fraction; Illustrate part of a whole as a fraction and as a %;	Convert simple fractions to % ;
	Relations	Read and write sentences involving greater than, less than and equal to using numbers up to 999;	Read, write and compare numbers using <, > and =; Identify, write and describe simple number patterns for factors and multiples;
	Measurement	Measure; length and height of objects; 2. use appropriate units in measurement above;	Measure; length, mass, area, perimeter, angle
OPERATIONS	Addition	Add up to three 1- to 2-digit whole numbers with multiple regroup; Solve simple problems on everyday use of addition;	Add up to three 1- to 4-digit whole numbers with and without regroup; Add proper fractions with 1- or 2-digit denominators that are equal or are multiples; Add decimal numbers with up to 2 decimal places; Know how to add '0' in both whole and decimal numbers Solve simple problems on everyday use of addition
	Subtraction	Subtract up to 2-digit from up to 3-digit whole numbers with and without regroup;	Subtract two 1- to 4-digit whole numbers with multiple regroup; Subtract proper fractions with 1- or 2-digit denominators that are equal or are multiples; Subtract decimal numbers with up to 2 decimal places Subtract numbers involving '0' in both whole and decimal numbers with 2 decimal places; Solve simple problems on everyday use of subtraction;
	Multiplication	Multiply 2-digit by 1-digit whole numbers with and without regroup,	Multiply up to 3-digit by 1- or 2-digit whole numbers with and without regroup;

SKILL COMPONENT	ELEMENTS	YEAR 4 INDICATORS (9 – 10 YRS)	YEAR 6 INDICATORS (11 – 12 YRS)
OPERATIONS	Division		Divide 2-digit whole numbers by 1-digit factor; Use order of operations to simplify expressions involving 2 operations; 3. solve simple problems on everyday use of the four operations
MONEY	Recognize and use money	Recognize the money value of items in shops, market, etc; Calculate costs of shopping with 2 items (in whole value only); Calculate changes from shopping;	Calculate total costs of shopping with 3 different items (include decimals); Calculate the change from shopping; Calculate the unit cost of items in shopping;
TIME	Time	Identify short/long hand or 1st/2nd number (digital) with hour/minute; Understand am/pm in relation to time of day; Tell time from clock face or diagram (but limit to o'clock, quarter past/to and half past);	Tell time from clock face or diagram; Calculate time difference from clock; Solve simple everyday problems on time and duration;
DATA	Data		Tally given sets of discrete data; Represent data on graph (bar orpicture); Interpret data on graph based on heights of bars (bar graph) and number of pictures (pictograph); Know and calculate the average of discrete data;



2016 PACIFIC ISLANDS LITERACY AND NUMERACY ASSESSMENT (PILNA)

NEW ZEALAND
FOREIGN AFFAIRS & TRADE
Aid Programme

Pacific Community
Communauté du Pacifique