Science Literacy in the International Baccalaureate Primary Years Programme (PYP): NAP-SL Outcomes

Final Report

10 March, 2014

Prepared by Associate Professor Coral Campbell, Dr. Gail Chittleborough, Dr. Wendy Jobling, Professor Russell Tytler & Dr. Brian Doig

School of Education, Deakin University

Contents

Executive Summary

With the substantial increase of International Baccalaureate (IB) schools across the world, the IB has recognised that support and future development of the organisation's programmes will require information relating to the impact and value of an IB education. As part of this agenda, the IB Research Department sought to understand the performance of Year 6 students, enrolled in the Primary Years Programme (PYP), against the Australian Proficiency Standards for Science Literacy as well results from the 2012 National Sample Assessment in Science Literacy.

This research project, by Deakin University, evaluated the science literacy of a sample of students undertaking the IB Primary Years Programme (PYP) in schools around Australia. This was accomplished by administering the National Sample Assessment in Science Literacy (NAP-SL) 2012 to children in ten schools, representing a mix of: state, urban, regional, government, and non-government schools. IB student tests were analysed and compared with the NAP-SL 2012 results and Australian Proficiency Standards. Comparisons were also made between male and female IB students, and IB students from government or non-government schools. State comparisons were made where possible. Please note that results marked ** rely on low numbers of students or samples.

Key Results emerging from the analysis indicate:

IB PYP students performed well in the NAP-SL test when compared with national results. For instance, the IB mean score result was 475.4089, whereas the national result for the NAP-SL was 401.1667. For Year 6 students, the anticipated national science proficiency Level is 3.2, which encompasses a score range of 393-523. This indicates that on average, IB students are placed towards the top of the Level 3.2 range, whilst the students in the national test were sitting close to the bottom of the range.

Other results include:

- The proficiency level of students in the IB PYP programme was generally higher than the national levels, with 83.3% of IB students at or above the suggested Year 6 Proficiency level of 3.2, compared with 51.4% of national students who undertook the NAP-SL test in 2012. The following table shows the percentage of IB students at each of the five science proficiency levels.

Distribution of Proficiency Level for Year 6			
Proficiency Level	Percentage of National sample(2012) in Proficiency level	IB Results for Practical and Objective test only	
		N=310	%
Level 4 and above	0.3	3	1.0
Level 3.3	9	70	22.6

Level 3.2	**42.1**	**185**	**59.7**
Level 3.1	**39.6**	**22**	**7.1**
Level 2 and below	**9**	**8**	**2.6**

- State comparisons could not be reliably made due to low numbers of test results in all states. However, with the data obtained, comparisons to State results indicate that IB PYP schools, in all states, gained higher test results overall than non-IB schools. See Table 11 for further details.

- Gender based proficiency comparisons indicate that IB students, both male and female, performed better on the NAP-SL than males and females in the national sample. Moreover, higher proportions of PYP males and female students achieved proficiency levels characterised as 'exemplary performance' than males and female students in the national sample.

- Students from government IB PYP schools generally out-performed students from non-government IB PYP schools. However, a caveat to this result is that there were just two government IB PYP schools in the study.

- Urban PYP students generally performed better than PYP students from regional schools. Mean scores for these groups are reported in Table 7, page 22.

- The small data set for girls' results meant it was not possible to make reliable comparisons between PYP boys and PYP girls. Although raw data suggests that PYP boys outperformed girls, this would need to be further investigated using larger data sets. In the actual test scores, PYP males' mean score was 29.1 out of a possible 51, whilst PYP females mean score was 27 out of a possible 51.

Introduction

The strategic importance of science is being highlighted at national and state levels (Chubb, 2012). The need for high quality science teaching and science education that engages students at all levels is crucial in developing a scientific literate population. Scientific literacy prepares future citizens for “interacting in a global environment needing to know how to learn, adapt, create, communicate, and how to interpret and use information critically”, and be able “to make personal decisions on the basis of a scientific view of the world” (National Curriculum Board, 2009, p. 4). As a result of scientific literacy skills being recognised as essential to understand the world around us, and to engage in discourses of and about science, the Australian Science Curriculum studied by Australian primary school students provides a strong focus on scientific literacy. In framing the development of the curriculum, two key documents articulating scientific literacy and its link to inquiry were used by the Australian Curriculum and Assessment Authority (ACARA): The review on *The status and quality of teaching and learning of science in Australian schools* (Goodrum, Hackling and Rennie 2001) and the monograph based on an important Australian Council for Educational Research (ACER) national conference: *Re-imagining Science Education* (Tytler 2007). Through the influence of these seminal documents, the three strands of the Australian Curriculum: *Science Understanding, Science Inquiry Skills, and Science as a Human Endeavor*, aim to support the development of student scientific literacy. In particular, the Science as a Human Endeavour dimension flags the importance of science and society links, consistent with the socio-scientific issue (SSI) focus advocated by many researchers (Tytler & Hobbs 2011).

With the substantial increase of International Baccalaureate (IB) schools across the world, the IB has recognised that support and future development of the organisation’s programmes will require information relating to the impact and value of an IB education. As part of this agenda, the IB Research Department sought to understand the performance of Year 6 students, enrolled in the Primary Years Programme (PYP), against the Australian Proficiency Standards for Science Literacy and national results of the 2012 National Sample Assessment in Science Literacy. The key research questions posed by the IB, which were subsequently investigated by Deakin University were:

1. *How do PYP students perform on the Australia National Sample Assessment in Science Literacy?*
2. *How do the results achieved by PYP students and schools compare with National and State results?*

The following report outlines the approach taken by the Deakin University team to addressing these questions. First we provide a background to the study, which includes an overview of the NAP-SL. Next, we outline the research design and report the results of the study. The report concludes by acknowledging study limitations, and offers several recommendations for IBs consideration.

Background to the Study

The place of science in the Australian Curriculum

> The Australian Curriculum sets out the core knowledge, understanding, skills and general capabilities important for all Australian students. It describes the learning entitlement of students as a foundation for their future learning, growth and active participation in the Australian community. It makes clear what all young Australians should learn as they progress through schooling.
> http://www.australiancurriculum.edu.au/

Within this framework, the 'Australian Curriculum: Science' provides a means for students to learn about the world around them. Science teaching and learning should reflect the multiple facets of science and serve a range of purposes. Science provides students with a way of answering their own questions about the biological, physical and technological world. This knowledge furnishes students with the basis for interacting in the many socio-scientific issues which affect their lives and for which their informed response is required. Students make sense of our world through exploring, making predictions and solving problems. Students develop an understanding of the processes of science and the important science concepts which underpin the 'big ideas' in the biological, chemical, physical, earth and space sciences. A strong aspect of the Australian science curriculum is the inclusion of the 'Science as a Human Endeavour' dimension which highlights the contribution of science to our culture and society and assists students to understand how they directly impact on local, national and global societies.

In undertaking science investigations, students develop a range of higher-order thinking skills. They are able to pose investigable questions and draw evidence-based conclusions using scientific methods. This is the basis of "scientific literacy" - having the science knowledge and skills, and the ability to apply them to investigate the natural world and make informed decisions in relation to their personal and public lives.

The Australian Curriculum: Science is based on six foundational ideas: patterns, order and organisation; form and function; stability and change; systems; scale and measurement; and matter and energy.

(Adapted from http://www.australiancurriculum.edu.au/Science/Rationale)

The IB PYP programme and the place of science

> The International Baccalaureate aims to develop inquiring, knowledgeable and caring young people who help to create a better and more peaceful world through intercultural understanding and respect.
> (International Baccalaureate Organization, 2007, pg. 2.)

To this end, the IB PYP programme attempts to develop students' capabilities and produce learners who are: inquirers; knowledgeable; thinkers: communicators; principled; open-minded; caring; risk-takers; balanced, and; reflective (IBO Learner Profile 2013). In the development of the PYP, the IB has considered the current research around international education and the implementation of thorough assessment regimes. The aim of their programmes is to "encourage students across the world to become active, compassionate and lifelong learners" (Making the PYP happen: A curriculum framework for international primary education, pg. 2.) The PYP supports an inquiry approach to learning whereby the students make sense of the world through active engagement with it. Students gather information, link it to experience and through analysis and synthesis create new meaning for themselves.

In the PYP science curriculum area, science is considered an exploration of the biological, chemical and physical world and the relationships between them. A knowledge of science is considered important and at an international level, "transcends the boundaries of gender, cultural, linguistic and national biases." (International Baccalaureate Organization, 2007, pg. 93). In this way, students develop an appreciation of the contribution of all peoples to the development of scientific understanding.

In the PYP, science learning is developed through a number of broad trans-disciplinary themes which are considered as 'organisers' for units of inquiry, although student-initiated learning is also a considered component of the programme. The knowledge element of science is explored through the strands of: living things, earth and space, materials and matter, and forces and energy. In addition, whilst the PYP programme supports the development of trans-disciplinary skills (thinking skills, social skills, communication skills, self-management skills and research skills), the science component also focuses on science-specific skills such as observation, using instruments and tools effectively, using scientific vocabulary , identifying a problem or issue, planning and executing an investigation, and making and testing predictions.

Within the PYP science curriculum, schools and teachers design and implement a programme which best suits the student body, individual cohorts and students.

Research Design

The stated aim of this study was to administer the National Sample Assessment in Science Literacy (NAP-SL) in PYP schools across Australia. In addition, after the marking of the assessments, the analysis of the results were compared to state and national results as well as national proficiency standards. Comparisons among IB students and schools were also made. This project involved the process of administration of test material, the analysis of results and the documentation of the results through mid, draft and final reports.

Research questions

A total of 337 IB Primary Years Programme students, attending government and non-government schools, completed the NAP-SL to assess PYP student performance in scientific literacy. In addition, a number of defined comparisons were undertaken. These include studying the national and state mean scores and comparing these to the PYP NAP-SL results. These activities were undertaken to address the following research questions:

How do PYP students perform on the Australia National Sample Assessment in Science Literacy?
How do the results achieved by PYP students and schools compare with National and State results?

The National Assessment Program – Science Literacy (NAP-SL)

The NAP-SL is one of several testing regimes developed by the Australian Curriculum, Assessment and Reporting Authority (ACARA). It is conducted every three years with a randomly selected sample of students from across Australia. The NAP-SL provides information on the scientific literacy of Grade 6 children. The public report (ACARA, 2013) provides provide "insight into the level of science-based knowledge, understandings and skills that our Year 6 students have developed".

The release material for NAP-SL supplied by ACARA on its website, which are freely available for schools to use to gauge student performance in relation to science literacy standards and national NAP-SL results consist of a Teacher's Instruction Guide (8pgs), Student Practice Questions (1pg), Student Workbook (26pgs), and a Practical Task (8pgs). These materials are clear and easy to understand. In addition, there are marking guidelines, analysis sheets and proficiency standards.

Proficiency standards

Proficiency standards were developed by the Australian Curriculum, Assessment and Reporting Authority (ACARA) in consultation with a range of people from various areas of education. Proficiency standards highlight what science skills children should be able to demonstrate at different levels. For grade six children, the target group for the NAP-SL test, the anticipated proficiency level is 3.2 (see: Appendix One for a table describing the proficiency levels), although the NAP-SL test covers five Proficiency Levels: Level 2, Level 3.1, Level 3.2, Level 3.3 and Level 4. According to ACARA, students whose results are located within a proficiency level are able to demonstrate the understandings and skills of that level. It would be expected that they would also possess the skills and understandings of the lower Proficiency levels.

The proficiency standards for scientific literacy have been based on three strands, compressed from the original 5 in PISA 2000 (OECD, 1999). These include:

- **Strand A** - Formulating and identifying investigable questions and hypothesis, planning investigations and collecting evidence.
- **Strand B** - Interpreting evidence and drawing conclusions from their own or others' data, critiquing the trustworthiness of evidence and claims made by others and communicating findings.
- **Strand C** - Using science understandings for describing and explain natural phenomena and for interpreting reports about phenomena.

In the 2012 ACARA NAP Science literacy testing, the following findings were detailed:

> The proficient standard was found to be equivalent to Level 3.2; that is, students achieving at Level 3.2 are considered to have a sound understanding of Year 6 science. Students at this level demonstrate considerably more skill and understanding than those performing at Levels 3.1 and below.
> Year 6 students who exceed the proficient standard (those who perform at Level 3.3 and above) demonstrate exemplary performance.
> (ACARA NAP SL 2012 School Release Materials.)

Data collection methods

Eleven schools were chosen from IB schools which have administered the PYP programme for at least three years[1]. The three year criterion was important so that the results can be interpreted as being the result of the PYP programme, not previous school teaching processes. Schools selected represent the IB national school population in terms of numbers in each state, whether government or non-government and whether urban or rural/regional.

In the first instance, schools were approached personally by a Deakin University academic staff member to explain the project and invite participation. This was followed up with the formal processes of providing documentation such as plain language statements that explain the project, consent forms and revocation of consent forms. Follow-up occurred after several days to arrange for test booklets and consent forms to be sent to the school.

The test booklets were printed versions of the ACARA materials, with one small variation. On the front page of the document, coding was used for the state, school, grade and

[1] Note – one school did not return test papers, effectively reducing school numbers to ten

student. All identifiers were removed. A student's age was requested in years and months. This was to ensure an accurate comparison with the ACARA test results.

The ACARA test materials consist of Teacher's Instruction Guide (8pgs), Student Practice Questions (1pg), Student Workbook (28pgs), and a practical task (8pgs). These materials, supplied in PDF format were printed off for the teachers and students involved. In addition, marking guidelines, analysis sheets and proficiency standards were provided. However, the recording sheet supplied was modified to allow Deakin academic staff to further analyse the types of errors or responses given by the students.

A moderation process was applied to the assessment of student scripts with all academic staff involved undertaking a discussion of the assessment items and the guidelines for assigning marks developed by ACARA for the tests, as a form of induction. When the first of the students' test items became available, the markers again met to moderate several tests and to ensure a consistent response to the application of the marking guidelines. In this way, coherent and uniform judgments were made across all assessment items that were as consistent as possible with the ACARA marking processes. While some test items are multiple-choice and do not require academic judgment, others are open questions and required a consistent approach in applying academic judgment. The use of the supplied marking guidelines, along with the induction meetings, allowed for accurate assessment and facilitated comparisons with the state and national results.

Sampling strategy

The sampling strategy was based on a number of factors. A proportional representative sample of 11 sites was used as the basis for the data collection. Sites were apportioned nationally to represent populations of students, and wherever possible included both non-government and government schools. Within the total number of IB schools in Australia, there is approximately 29% government and 71% non-government (independent) PYP schools. A table was drawn up to indicate the number of IB schools in each state of Australia. The selection of schools was then undertaken using the requirements of the tender for particular comparisons. The research needed representative schools relating to government/non-government, urban and regional, and from within each state.

Apart from comparing the total student body to the results of the National standards, comparisons of males to females, government to non-government and state to state, were attempted, with the knowledge that numbers are relatively small for a full statistical analysis in some cases, and with a greater number of schools from the non-government sector. A final proportion of 8 non-government to 3 government sites was used to represent a 73:27 ratio. With the loss of one government school, this ratio was altered to 80:20.

To address the requirement of comparison with state and national results, it was important to have schools selected from each state. IB schools are unevenly spread across states, with two of the states, Victoria and South Australia, having between them 61 of the 86 IB schools nationally. The inclusion of both a government school and a non-government school from each of South Australia (SA) and Victoria was intended to allow some sector comparisons to be made. Although a sampling of 3 schools were achieved for SA, the overall number of students involved was low so that, although state comparisons were undertaken, the validity of the results needs to be defined from the standpoint of insufficient numbers for accurate comparisons.

Table 1: *Anticipated Sampling U=Urban, R= Rural/Regional*

Spread of IB PYP schools in Australia	**Population total**	**Sample size req.**	**Govt**	**Sample govt**	**Non-govt**	**Sample Non-govt**	**Number of students total**
ACT (4 U)	4	1u	2	0	2	1	40-50
NSW (7 U, 1 R)	8	1u	0	0	8	1	40-50
QLD (4 U, 4 rural)	8	1U +1R	1	0	7	1	40-50
SA (20U, 7R)	27	3U +1R	7	1	20	3	160-200
TAS (1U)	1		0	0	1	0	0
VIC (33U, 1R)	34	3U+1Regional	15	2	19	2	160-200
WA (4U)	4	1	0	0	4	1	40-50
Total	**86**	**12**		**3**		**9**	**480-600**

Two grades were selected from each school which allowed for a student cohort of between 40-50 children at each site. It was assumed that approximately 50% of the students (at each site) would be male, and 50% female. This would be a statistically viable number to allow cross-school comparisons as well as male/female comparisons within schools and across the total sample. Unfortunately, in the final breakdown of numbers, we did not achieve an adequate proportion of females (26%) when compared with the number of males (74%).

In considering the sampling regime, we were cognisant that there may be a range of other variables affecting the results, in particular the socio-economic status of the child and school. The national NAP-SL data from both the School Release Materials and the public report, provided no detailed breakdown of results in socio-economic profile, so it was not possible to undertake a 'like school' comparison.

Development and Implementation of the Primary Years Programme NAP-SL Study

Deakin ethics and jurisdictional approvals

Deakin Ethics and most jurisdictional approvals were gained without undue problems. However:

- The Australian Capital Territory Education Training Directorate did not give consent to the research and so we had to exclude ACT government schools from the study, although we were able to include a non-government ACT school in the selection; and
- Approval from a Catholic diocese was sought in order to comply with CEO approval process and the recruitment of a Catholic school in Adelaide. Some delays in this were caused by difficulties in gaining information from Catholic Dioceses concerning details of the process.

Release and printing of ACARA School Release Materials

There was substantial delay in ACARA releasing 2012 results, with the materials being supplied to Deakin University, along with permission to use their materials, on 19th August instead of the expected mid-July release time. Printing of all materials was completed before the end of August and packages were sent to schools as soon as principals and/or senior staff agree to participate. The 2012 materials were released publically in November 2013.

School recruitment

The time it took for recruitment of schools to the project was longer than anticipated as schools often followed a process of internal discussion and it often took up to 3 weeks for schools to respond to the invitation to participate. One school responded positively to the invitation but took a further two months before undertaking the tests and returning them – even with timely reminders. However, schools seemed genuinely interested in the study and those that declined to participate generally did so for one of the following reasons:

- Competing demands (e.g. IB Exhibitions, other IB funded research being conducted concurrently)
- Grade 6 numbers were too small (e.g. 1-2 students).

It should be noted that the competing demands of Year 6 Exhibitions were significant factors in delaying the progress of the project. At least 35 schools were invited to participate but over 70% declined. In some states, contact was made with every available school in an attempt to find enough students for the test samples. We eventually recruited 11 schools - 8 urban (3 govt /5 non-govt) and 3 regional (all non-govt), although one school failed to return any tests, despite reminders. The team used a letter of support from IBO to assist the process of recruitment. This necessity arose from one school questioning our authority to undertake the research and refusing to accept the Deakin documentation as evidence. Tests from schools were still being returned as late as 15 December, making analysis and reporting a very difficult task to complete within the time constraints of the research contract. IBO Research Department offered an extension on the submission of the report. Table 2 provides details of the participating schools including location and total numbers of test returns.

Table 2: *A breakdown of the ten participating schools and their characteristics*

School	**State**	**Urban**	**Regional**	**Govt**	**Non-govt**	**Grd 6 students**	**Test Returns**
A	VIC	✓			✓	56	21
B	SA	✓		✓		27	20
C	WA	✓			✓	80	56
D	QLD		✓		✓	20	21
E	ACT	✓			✓	50	43
F	SA		✓		✓	10	3
G	NSW	✓			✓	76	74
H	VIC	✓		✓		74	48
J	SA		✓		✓	10	10
K	SA	✓			✓	78	41
	TOTAL					**550**	**337**

Results – PYP NAP-SL Outcomes and Proficiency Levels

All IB students' NAP-SL tests were marked and results were compared in a number of ways. Initially, for each test item, all the IB student results were compared to the national results for that test item. Figure 1 shows the proficiency level accorded to test items and the number and proportion of PYP students who answered questions correctly. The column on the right of the table provides the national results for each question.

Figure 1: Comparison of all questions, IB: ACARA NAP-SL 2012 sample

OBJECTIVE ASSESSMENT				
Sample		**337**	**337**	5%
Item number	**Proficiency Level**	**Number of students correct**	**IB Science National % Correct**	**ACARA National % correct**
Q1	≤ 2	309	91.7	86.6
Q2	3.2	247	73.3	61.5
Q3	3.1	295	87.5	75.2
Q4	3.1	272	80.7	75.4
Q5	3.2	183	54.3	41
Q6	3.3	158	46.9	33.6
Q7	3.1	302	89.6	81.2
Q8	3.2	212	62.9	43.7
Q9	3.1	300	89.0	81.1
Q10	3.2	209	62.0	44.5
Q11	3.3	193	57.3	33.9
Q12	3.2	253	75.1	58.9
Q13	3.3	133	39.5	32.3
Q14	3.2	224	66.5	55.9
Q15 (1 or 2 marks)	3.2; ≥ 4	165; 59	49.0; 17.5	46.5; 14.4
Q16	3.1	267	79.2	64.7
Q17	3.2	201	59.6	47.3
Q18	3.3	181	53.7	34.6
Q19	3.2	237	70.3	53.4
Q20	3.2	193	57.3	55.1
Q21	3.3	155	46.0	32.4
Q22	3.2	161	47.8	37.8
Q23	3.2	161	47.8	62.1
Q24	≥ 4	28	8.3	3.7
OBJECTIVE				

ASSESSMENT continued…				
Sample		**337**	**337**	**5%**
Item number	**Proficiency Level**	**Number of students correct**	**IB Science National % Correct**	**ACARA National % correct**
Q25	3.2	220	65.3	50.9
Q26	3.2	223	66.2	48.3
Q27	3.2	254	75.4	53.2
Q28	3.2	216	64.1	45.9
Q29	3.3	105	31.2	28.9
Q30	3.2	169	50.1	42.8
Q31	3.2	176	52.2	43.6
Q32	3.3	114	33.8	25.6
Q33	≤ 2	314	93.2	87.3
Q34	3.2	123	36.5	39.8
Q35	3.3	190	56.4	38.3
Q36	≥ 4	67	19.9	14.3
Q37	≥ 4	40	11.9	8.9
Q38	3.3	126	37.4	25.9
Q39 (1 or 2 marks)	≥ 4;≥ 4	85; 4	25.2 ; 1.2	16.7; 0.7
PRACTICAL TASK		310	**310**	
Item number	**Proficiency Level**	**Number of students correct**	**IB Science National % Correct**	**National % correct**
P1	3.1	268	86.5	83.9
P2	≥ 4	146	47.1	6.8
P3	3.2	240	77.4	59.4
P4	3.1	266	85.8	71.3
P5	3.2	187	60.3	44.4
P6	≥ 4	67	21.6	5.4
P7	3.1	295	95.2	83.8
P8	3.1	294	94.8	82.7
P9	3.2	226	72.9	50
P10	≥ 4	74	23.9	3.5

As can be seen in Figure 1, for each and every question in the NAP-SL test, except for question 23, the students undertaking the PYP in IB schools out-performed the national average. Question 23 deals with evaporating liquids and the correct response indicates a student's awareness of the meaning of 'fair testing' by identifying one factor that should be kept the same in an experiment. For this question, three IB PYP schools did outperform the national result, but the overall result was that this question was poorly answered by PYP students.

Figure 1 also illustrates that for questions at the grade six proficiency level (3.2) or lower, the percentage of students getting correct answers was high, as would be expected. As the questions become harder, the number of students answering correctly, is reduced. In general, however, IB PYP students performed better on the more advanced level questions. Results for these are shown in Table 3.

Table 3: *IB students' performance on advanced level questions*

Item number	Proficiency Level	Number of students correct (n=337, P-n=310)	IB Science National % Correct	National % correct
Q24	≥ 4	28	8.3	3.7
Q36	≥ 4	67	19.9	14.3
Q37	≥ 4	40	11.9	8.9
P6	≥ 4	67	21.6	5.4
P10	≥ 4	74	23.9	3.5

PYP Science Literacy Proficiency Levels

As previously noted, all students were assessed against the five proficiency levels 2, 3.1, 3.2, 3.3, 4.0, to determine how they performed. Appendix Three provides an overview of the Proficiency Levels and the skills and understanding required for each strand.

Table 4 and Table 5 show a comparison of the profile of numbers of students performing at different proficiency levels for the IB schools compared to schools nationally. Whilst IB results were available for each separate test (objective test only and the objective and practical tests combined), the national results supplied in the public report were not reported separately. Each table therefore contains the reported national results which cannot be separated, and thus are the same for both tables.

Table 4 shows that 77.8% of PYP students performed at or above the targeted proficiency standard of 3.2 on the objective test, with only 22.2% below. For the national sample, the result was 51.4% above and 48.6% below.

Table 4: *A comparison of the profile of proficiency for IB compared to schools nationally for the objective test.*

Distribution of Proficiency Level for Year 6			
Proficiency Level	**Percentage of National sample(2012) in Proficiency level**	**IB Results for Objective test only**	
		N=337	**%**
Level 4 and above	**0.3**	**3**	**0.9**
Level 3.3	**9**	**70**	**20.8**
Level 3.2	**42.1**	**189**	**56.1**
Level 3.1	**39.6**	**63**	**18.7**
Level 2 and below	**9**	**12**	**3.6**

Table 5, below, provides the same result for the objective and practical tests combined, and indicates 83.3% of students undertaking the IB PYP were at or above the suggested Grade 6 proficiency level of 3.2. This compares favourably with the national sample where 51.4% of children were at or above the 3.2 level and 48.6% were below.

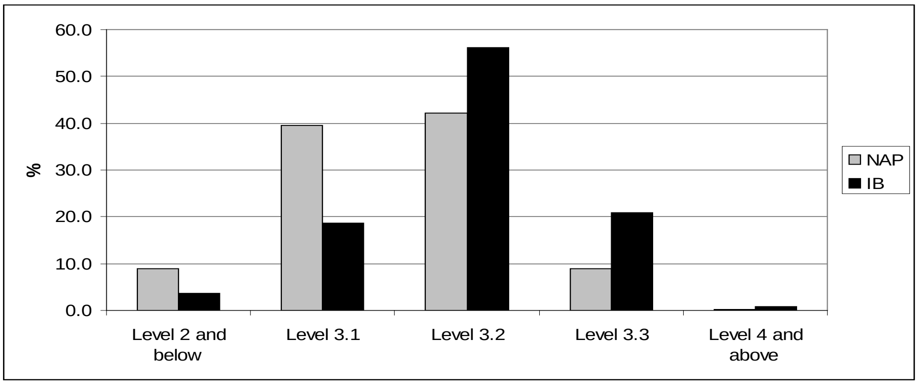
Table 5: *A comparison of the profile of proficiency for IB compared to schools nationally for the objective and practical tests combined.*

Distribution of Proficiency Level for Year 6			
Proficiency Level	**Percentage of National sample(2012) in Proficiency level**	**IB Results for Practical and Objective test only**	
		N=310	**%**
Level 4 and above	**0.3**	**3**	**1.0**

Level 3.3	**9**	**70**	**22.6**
Level 3.2	**42.1**	**185**	**59.7**
Level 3.1	**39.6**	**22**	**7.1**
Level 2 and below	**9**	**8**	**2.6**

A further notable result in Table 5 is that 23.6% of PYP students performed at proficiency level 3.3 or above, compared with 9.3% for the national sample. According to ACARA, year 6 students who "exceed the proficient standard (of 3.2) demonstrate exemplary performance" (ACARA, 2010:48). Figure 2 shows in a graphical form how IB students are clearly performing differently to the National sample. Although Figure 2 points to higher science proficiency levels among IB students in science, as will be discussed in a later section of the report, the result is not even across non-government and government schools. These discrepancies need to be investigated further, and in more detail. A caveat for the results presented here is that some states were represented by a single school, which may skew the results for the entire sample. However, this being said, these results provide clear evidence for the benefits of the IB programme in science by the end of primary school.

Figure 2: *Proportions of students performing at curriculum Level*



Lastly, Table 6 identifies that PYP students also performed higher than national results when proficiency levels for males and females were compared. Nationally, approximately 51% of female students performed at or above a proficiency level of 3.2; however, the figure was 79% for female PYP students. Meanwhile, 80% of male PYP students performed at or above a proficiency level of 3.2, whereas 52% of males nationally achieved or exceeded a proficiency level of 3.2. Much higher proportions of both male (25.1%) and female (16%) PYP students performed at a proficiency level 3.3 or above, when compared with national male (9.8%) and female (9.2%) NAP-SL results. Appendix Four provides a breakdown of IB and National NAP-SL gender results for Australian States and Territories.

Table 6: *Gender results by Proficiency levels*

Percentage of national sample(2012) in proficiency Level	Percentage of national sample(2012) in proficiency Level	Proficiency level	Percentage of IB students in proficiency level	Percentage of IB students in proficiency level
Male	**Female**		**Male**	**Female**
0.4	0.2	Level 4 and above	1.3	0
9.4	9	Level 3.3	23.8	16
41.9	42.30	Level 3.2	55.3	63
34.4	41.00	Level 3.1	17	18
9.9	8%	Level 2 and below	2.6	3

Further Results from the Analysis of PYP Science Literacy Data

To compare results among PYP students and schools, the raw scores from the sample of 337 PYP students were analysed using a Masters Partial Credit Model (Masters, 1982) a member of the Rasch (Rasch, 1960) family of Item Response Theory (IRT) models. The Quest (Adams & Khoo, 1996) was used to perform the Masters Partial Credit Model analysis. This analysis provides information about both the students and the items to which they responded. The Wright Map below (Figure 3) shows the details in a graphical form.

The Rasch analysis has a unique characteristic in that both the students and the items are placed on the same scale. In effect, this means that it is possible to estimate the likelihood of a student with a particular scale score responding correctly to an item of any difficulty. For example, a student with a scale score of 1 logit has a likelihood of responding correctly to Cognitive items 34 and 38 (see below), but is more likely to respond correctly to items below that point on the scale. In fact, as the items become lower on the scale than the student’s position, the likelihood of a correct response from this student increases. In an opposite manner, items on the scale above 1 logit are likely to be too difficult for that student, and the likelihood of a correct response decreases as the distance above their position on the scale increases.

In a Wright map there is a scale in logits on the left-hand side and a vertical line in the centre dividing the Map into two columns. The left-hand column shows the distribution of students along the logit scale (where an X denotes 2 students in this case). The students are ordered from the least proficient at the bottom up to the most proficient at the top. On the right-hand side of the Wright map the items are ordered from the easiest at the bottom to the most difficult at the top. The items are described by an alpha-numeric code as follows: the letters C and P identify an item as a cognitive or practical item respectively, the number indicates the item number in the assessment. There is a suggested proficiency level for each item in the NAP report and this should be read in conjunction with Figure 3. For example, C11 is item 11 in the cognitive assessment: it is suggested that this item assesses proficiency level 3.3. In two cases, items had two scores, or response categories, and these are denoted by a 1 or 2 after a period (.). For example, item C39, the most difficult item, is shown as C39.2. All this means is that for cognitive item 39, to score at the second level of response is very difficult, and is classified as being at proficiency level at, or above, four.

As expected most of the difficult items, representing a proficiency level of 4 are at the upper end of the logit scale, and the easier items, representing a proficiency level of 3.2 or 3.1, are at the lower end of the scale. An unexpected result shown by the Wright Map is item P6, which has a suggested level of 2, but is as difficult as other cognitive and practical items whose proficiency level is 4 or above. This suggests either, that the item has been misclassified, or that the students in this sample found this item particularly difficult, which is unlikely.

Also the Wright Map makes it clear that the items, as a group, are easy for the majority of students. This can be seen in the Map as the average of the students is above the zero logit point of the scale, which is the average scale difficulty of the item difficulties.

The Rasch analysis provided an ability estimate for all students on all items answered by the student. Some students answered both the Cognitive and the Practical items, while others answered only the Cognitive. All sub-group analyses used item estimates anchored on all student responses and thus are all on the same scale. Results for the combined assessments and sub-groups are described below in Table 7.

Figure 3: *Wright Map of IB PYP students and NAP items*

```
-----------MORE PROFICIENT------------------HARDER ITEMS-----
  5.0                            |COGNITIVE       PRACTICAL
                                 |
                                 |
                                 |
                                 |
                                 |    C39.2
                                 |
  4.0                            |
                                 |
                             X   |
                                 |
                                 |
                                 |
                                 |
  3.0                            |    C24
                                 |
                             X   |    C37
                             X   |                      P10
                                 |
                                 |
                            XX   |    C36
  2.0                     XXXX   |    C15.2
                         XXXXX   |                      P6
                         XXXXX   |
                           XXX   |    C39.1
                     XXXXXXXXX   |    C29
                          XXXX   |    C32
  1.0                   XXXXXX   |    C34   C38
          XXXXXXXXXXXXXXXXXXXX   |    C13
                    XXXXXXXXXX   |    C6                P2
                 XXXXXXXXXXXXX   |    C21
                       XXXXXXX   |    C23
              XXXXXXXXXXXXXXXX   |    C5    C30   C31
                        XXXXXX   |    C18   C35
   .0                 XXXXXXXX   |    C11   C20   C22
                  XXXXXXXXXXXX   |    C10   C17         P5
                       XXXXXXX   |    C8    C15.1 C25   C28
                      XXXXXXXX   |    C14   C26
                         XXXXX   |    C19               P9
                      XXXXXXXX   |    C2
                          XXXX   |    C12   C27         P3
 -1.0                      XXX   |
                            XX   |    C4   C16          P1
                            XX   |
                             X   |                      P4
                             X   |
                            XX   |    C3
                                 |
 -2.0                        X   |    C7   C9
                             X   |
                                 |                  P7   P8
                                 |    C1   C33
                                 |
                                 |
                                 |
 -3.0      LESS PROFICIENT       |        EASIER ITEMS
--------------------------------------------------------------
  Each X represents    2 students
==============================================================
```

Comparing PYP Results by Locality, School Type and Gender

The overall scale score results for all PYP students in this study, on all items, are shown in Table 7. These results have been summarized for various sub-groups as well as the totals for each sub-group of schools (i.e. government and non-government) and states.

Looking at the overall results, it can be seen that the mean for all PYP students (0.42 logits) is somewhat over-shadowed by the large Standard Deviation (0.88 logits). However, some of the sub-groups of schools offer some interesting results. For example, in Victoria, the students in 2 Government schools out-perform those in non-government schools, a situation that is reversed in South Australia.

Table 7: *Means and Standard Deviations for various totals and sub-groups*

		State	Type	Region	School	Name
ACT			**Non-government**	**Urban**		
	Mean	0.67	0.67	0.67	0.67	School E
	SD	0.91	0.91	0.91	0.91	
New South Wales	**Mean**	0.41	0.41	0.41	0.41	School G
	SD	0.91	0.91	0.91	0.91	
Queensland	**Mean**	0.38	0.38	0.38	0.38	School D
	SD	0.82	0.82	0.82	0.82	
South Australia			**Non-government**	**Urban**		
	Mean	0.11	0.26	0.28	-0.37	School K
	SD	0.79	0.82	0.74	0.27	
	Mean				0.30	School F
	SD				0.83	
			Government	**Rural**		
	Mean		-0.09	-0.55	-0.73	School J
	SD		0.70	0.65	0.56	
	Mean				0.23	School B
	SD				0.54	
Victoria			**Non-government**	**Urban**		
	Mean	0.37	0.14	0.37	0.14	School A
	SD	0.91	0.71	0.91	0.71	
			Government			
	Mean		0.47		0.47	School H
	SD		0.97		0.97	
Western Australia			**Non-government**	**Urban**		
	Mean	0.70	0.70	0.70	0.70	School C
	SD	0.77	0.77	0.77	0.77	
Study Total	**Mean**	0.42				
	SD	0.88				
Independent Total	**Mean**		0.37			
	SD		0.84			
Government Total	**Mean**		0.58			
	SD		0.96			
Urban Total	**Mean**			0.60		
	SD			0.80		
Regional Total	**Mean**			0.40		
	SD			0.88		

Further analysis of the results was achieved by applying Cohen's Effect Size which is a measure of the strength of a phenomenon – in this case, student results in the NAP-SL test in various categories. The effect size is calculated from statistical data (Means, and Standard deviation) and is a statistic that describes the estimated magnitude of a relationship. The Cohen's Effect Size measure does not imply or attempt to relate how the data reflects a true relationship in the population. The Cohen's Effect Size is calculated by finding the difference between the two means of the data and dividing this figure by the pooled standard deviation (see Table 8).

Table 8: *Interpretation of Cohen's Effect Size*

Cohen's Effect Size *d*	**Descriptor**	**Value**
	Low effect	0.0-0.15
	Medium effect	0.15-0.4
	High effect	0.4-1.2

Government and non-government PYP school comparisons

While Government PYP school students performed better than non-government PYP school students, (Mean 0.58; SD 0.96, and Mean 0.37; SD 0.84 respectively), both groups retain a very large Standard Deviation, indicating wide variation in attainment in different schools. Cohen's Effect Size was calculated as d = 0.24, which indicates that PYP students in Government schools are achieving better results than PYP students in non-government schools for the aspects of science assessed by the NAP, at a medium confidence level.

Urban and rural/regional PYP school comparisons

Turning to differences between urban and regional students, we find that urban students and regional students (Mean 0.60, SD 0.80; Mean 0.40, SD 0.88 respectively) differ marginally. A Cohen's Effect Size for these two groups of students (d = 0.25) suggests that children from urban schools are achieving moderately better results when compared to students from rural/regional schools.

Girls' and boys' IB comparisons

Table 9 provides details for the performance of both sexes on the combined Cognitive and Practical assessments. The relatively small number of girls, although compensated for by using an un-biased version of Cohen's d, makes the drawing of general conclusions problematic.

Table 9: *Performance of Boys and Girls*

	Mean	SD	N
Boys	0.48	0.89	244
Girls	0.27	0.84	88

As can be seen in Table 9, within the IB sample both sexes were represented, but there are far fewer girls than boys. Table 10 shows the mean scores and standard deviations of both groups.

Table 10: *Proficiency scores between sexes*

Assessment		Mean	Standard Deviation	N
Cognitive	Males	22.8	6.37	245
	Females	21.9	6.04	88
Practical	Males	6.3	2.15	245
	Females	5.1	2.71	88
Both	Males	29.1	7.67	245
	Females	27	7.53	88

The difference between the Means for both the Cognitive and Practical items is merely one score point, and the Standard Deviations are similarly very close. Effect sizes were calculated for each of Cognitive assessment, Practical assessment, and the combined assessment. The results were d = 0.143, d = 0.518, and d = 0.274 respectively. As stated above, the relatively small number of girls makes strong inferences unwarranted, but there would appear to be some support for drawing the conclusion that the Practical assessment was less well done by the girls as a group. The main contributing factor here is that the girl's Mean score is more than one score point lower than the boy's Mean, coupled with the girls' Standard Deviation being larger than that of the boys. This may be a point worth revisiting in a larger study, and if these results are confirmed, an investigation into girls' attitudes to, and proficiency in, practical assessment in science should be conducted.

State Comparisons with National data

To provide a State vs. National comparison, the two states with the largest number of participating schools were chosen to provide a better picture of the spread of results across the state. However, the data were compromised by the number of tests actually returned from the schools, essentially limiting the conclusions which can be drawn from the data. However, the data confirm the overall results stated earlier, that students undertaking IB PYP performed better than students involved in the National study. Looking at the State results below (Table 11), there is the inference that the IB PYP students, at State level, are achieving higher results than the State results. What should also be noted is that the between school variance is high, even when considering two schools in one state. The variation in IB test results in both Victoria and South Australia would be considered significant.

Table 11: *State comparisons with National data*

State	**IB School**	**Number of test results**	**IB Mean Score**	**category**	**National results/ state**
Victoria	A	21	460.75	U/NG	393 (Range 237-544)
	H	48	476.75	U/G	
SA	K	41	460.82	U/NG	392 (Range 234-542)
	F	3	Insufficient data	R/NG	
	J	10	388.9 **	R/NG	
	B	20	586.75	U/G	
WA	C	56	494.30	U/NG	406
Queensland	D	21	459.62	R/NG	392 (Range 236-544)
ACT	E	43	494.83	U/NG	429 (Range 269-580)
NSW	G	74	455.96	U/NG	395 (Range 231-560)

** denotes that the practical test results not included in the IB Mean score.

Conclusion

This report has outlined the aims and objectives of the Science Literacy in the IB Primary Years Programme study, the selection of sample schools, how children's scientific literacy was evaluated, the analyses of PYP schools' performance and the comparison of IB PYP school performance to state and national results.

The sampling strategy was based on a number of factors. A proportional representative sample of 12 sites was to be used as the basis for the data collection. Sites were apportioned nationally to represent populations of students, and wherever possible include both non-government and government schools. To address the requirement of comparison with state and national results, it was important to have schools selected from each state. In two of the states, Victoria and South Australia, the inclusion of both a state school and a non-government school allowed some sector comparisons to be made. Unfortunately for a number of reasons indicated earlier in this report, we were unable to achieve the breadth or depth of assessment data from schools from around Australia that we set out to achieve.

In considering the sampling regime, we were cognisant that there may be a range of other variables affecting the results. For example, we were aware that other aspects such as the size of the school, the numbers of children in each classroom, the distributions of girls and boys within and across schools and the socio-economic status of the child and school, could impact on the results. However, the national NAP-SL data from both the School Release Materials and the public report, provided no detailed breakdown of results in regards to any of these elements. Therefore, it was not possible to undertake a 'like school' comparison. This was not a requirement of the tender, which requested an overall comparison, and targeted comparisons by state, gender and rurality. We acknowledge that in arriving at the conclusions written in this report, factors outside the research methodology may have impacted on the outcomes of the research.

As indicated in the analysis of data, key results emerging from the analysis indicate:

- IB PYP students performed well in the NAP-SL test when compared with the national results;
- The proficiency level of students in the IB PYP programme is generally higher than the national levels;
- The gender proficiency comparisons also indicate that IB students, both male and female, achieve higher proportions at the higher proficiency levels than males and females national results;
- Comparisons to National State results suggest that IB PYP schools in all states, gained higher test results overall;

- Students from government IB PYP schools generally out-performed students from non-government IB PYP schools;
- Urban PYP students generally performed better than students from regional PYP schools;
- The small data set for girls' results meant it was not possible to make reliable comparisons between PYP boys and girls – further investigations using larger data sets is required.

Even within the sampling limitations and with the knowledge that the overall numbers of test results in each school was low, the data are consistently indicating that students in IB PYP settings achieved higher results in the 2012 National Sample Assessment in Science Literacy when compared with students who sat the national test in 2012. Some results were unexpected and, although based on limited data sets, statistically significant. For example, girls and boys comparison did not necessarily match those of the national comparisons where there was no significant difference in results (NAP-SL Public Report, 2013). This highlights the comment made earlier that for a more robust result, a larger study would be required.

In concluding, it can be said that the results of the research suggest that IB PYP school students are achieving well in terms of science literacy when compared with the national results.

Recommendations

It is recommended that further studies are undertaken to confirm some of the interpretations of this study. One such area of investigation is to look at whether the particular IB approach to inquiry was responsible for these enhanced results. Would it be possible to study if there is a difference in the culture of teaching science that the IB stimulates? Additionally, there would appear to be a need to investigate the support needs of rural schools given the gap in performance. This may be related to the broader issue of professional support for teachers in rural areas. Finally, one test result was unusual (question 23 appeared to be poorly answered by IB PYP students) and may need further exploration to determine the cause.

References

http://www.nap.edu.au/verve/_resources/2009_SL_School_release_materials_.pdf
ACARA NAP SL (2012) School Release Materials – provided to Deakin University on 19 August 2013.

ACARA, NAP SL (2013) Public Report Accessed Nov 2013 http://www.nap.edu.au/search-results.html?query=science

Adams, R. J., & Khoo, S-T. (1996). Quest: The interactive test analysis system, Version 2. Camberwell: ACER

Chubb (2012).*Mathematics, Engineering and Science in the National Interest.* Canberra: Office of the Chief Scientist. Viewed February 2014, http://www.chiefscientist.gov.au/wp-content/uploads/Office-of-the-Chief-Scientist-MES-Report-8-May-2012.pdf

DeFife, J. (2009). Effect Size Calculator: Emory University.

OECD Programme for International Student Assessment (1999) Measuring Student knowledge and skills: a new framework for assessment. Paris. OECD

National Curriculum Board, (2009) The Shape of the National Curriculum: A Proposal for Discussion. Accessed 11 October 2013 http://www.ncb.org.au/verve/_resources/ The_Shape_of_the_National _Curriculum_paper.pdf

Goodrum, D. Hackling, M. & Rennie, L. (2001) The status and quality of teaching and learning of science in Australian schools: a research report. The Department of Education, Training and Youth Affairs. Accessed 11 October 2013http://www.dest.gov.au/sectors/school_education/publications_resources/science_in_australian_schools

International Baccalaureate Organization, (2007) Primary Years Programme, Making the PYP happen: A curriculum framework for international primary education

International Baccalaureate Organization, (2008) Primary Years Programme Science scope and sequence.

International Baccalaureate Organization, (2011) Science Across the IB Continuum

International Baccalaureate Organization, (2013), IB Learner Profile http://www.ibo.org/programmes/profile/

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

NAP-SL Public Report http://www.nap.edu.au/verve/_resources/NAP-SL_2012_Public_Report. Accessed December 2013

NAP-SL School Release Materials http://www.nap.edu.au/verve/_resources/NAP-SL_2012_School_Release_Materials.pdf

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests (Expanded ed.)*. Chicago: University of Chicago Press.

Tytler, R. (2007) *Re-imagining Science Education*. Australian Council *for* Educational Research, ACER Press. http://www.acer.edu.au/documents/

Tytler, R. & Hobbs, L. (2011). The Australian Science Curriculum*, Primary and Middle Years Educator*, vol. 9, no. 2, pp. 3-10.

Appendices

Appendix One

Description of skills assessed at each proficiency level

Proficiency Level	Descriptor: a student at this level may display skills like
Level 2 and below	Makes measurements or comparisons involving information or stimulus in a familiar context. Identifies simple patterns in the data and/or interprets a data set containing some interrelated elements. Makes a choice for a situation based on first-hand concrete experience, requiring the application of limited knowledge.
Level 3.1	Makes simple standard measurements and records data as descriptions. Interprets simple data set requiring an element of comparison. Selects appropriate reason to explain reported observation related to personal experience.
Level 3.2	Collates and compares data set of collected information. Selects experimental design that respresents a fair test. Interprets data and identifies patterns in – and/or relationships between – elements of the data. Interprets information in a contextualised report by application of relevant science knowledge.
Level 3.3	Demonstrates an awareness of the principles of conducting an experiment and taking into account variables to be changed and/or measured. Extrapolates from an observed pattern to describe an expected outcome or event. Applies knowledge of relationship to explain a reported phenomenon.
Level 4 and above	When provided with an experimental design involving multiple variables, can identify the questions being investigated. Conclusions summarise and explain the patterns in the data in the form of a rule and are consistent with the data. Explains interactions that have been observed in terms of an abstract science concept.

Appendix Two

Strand Descriptors for each proficiency level

Proficiency level	Strand A Formulating and identifying investigable questions and hypothesis, planning investigations and collecting evidence.	Strand B Interpreting evidence and drawing conclusion from their own or others' data, critiquing the trustworthiness of evidence and claims made by others and communicating findings.	Strand C Using science understandings for describing and explain natural phenomena and for interpreting reports about phenomena.
Level 2 and below (scaled score ≤ 262)	Given a question in a familiar context, identifies that one variable/factor is to be changed (but does not necessarily use the term variable to describe the changed variable. Demonstrates intuitive level of awareness of fair testing. Observes and describes or makes non standard measurements and limited records of data.	Makes comparisons between objects or events observe. Compares aspects of data in a simple supplied table of results. Can complete simple tables and bar graphs given table column headings or prepared axes.	Describes changes to, differences between or properties of objects or events that have been experienced or reported
Level 3.1 (scaled score 262–393)	Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair testing and appreciates scientific meaning of 'fair testing'. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	Displays data as tables or constructs bar graphs when given the variables for each axis. Identifies and summarises patterns in science data in the form of a rule. Recognises the need for improvement to the method. Applies the rule by extrapolating and predicting.	Describes the relationships between individual events (including cause and effect relationships) that have been experienced or reported. Can generalise and apply the rule by predicting future events.
Level 3.2 scaled score 393–523)	Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair testing and appreciates scientific meaning of 'fair testing'. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	Displays data as tables or constructs bar graphs when given the variables for each axis. Identifies and summarises patterns in science data in the form of a rule. Recognises the need for improvement to the method. Applies the rule by extrapolating and predicting.	Describes the relationships between individual events (including cause and effect relationships) that have been experienced or reported. Can generalise and apply the rule by predicting future events.
Level 3.3 (scaled score 523–653) ** Note Level 3.2	Formulates simple scientific questions for testing and makes predictions. Demonstrates awareness of the need for fair	Displays data as tables or constructs bar graphs when given the variables for each axis. Identifies and summarises	Describes the relationships between individual events (including cause and effect relationships) that have been

and 3.3 descriptors are the same until the next level down.	testing and appreciates scientific meaning of ‘fair testing’. Identifies variable to be changed and/or measured but does not indicate variables to be controlled. Makes simple standard measurements. Records data as tables, diagrams or descriptions.	patterns in science data in the form of a rule. Recognises the need for improvement to the method. Applies the rule by extrapolating and predicting.	experienced or reported. Can generalise and apply the rule by predicting future events.
Level 4 and above (scaled score > 653)	Formulates scientific questions, identifies the variable to be changed, the variable to be measured and in addition identifies at least one variable to be controlled. Uses repeated trials or replicates. Collects and records data involving two or more variables	Calculates averages from repeat trials or replicates, plots line graphs where appropriate. Interprets data from line graph or bar graph. Conclusions summarise and explain the patterns in the science data. Able to make general suggestions for improving an investigation (e.g. make more measurements).	Explains interactions, processes or effects that have been experienced or reported, in terms of a non-observable property or abstract science concept.

Appendix Three

Extract from NAP SL Public Report (2013) Page xxi

Table ES.3 Percentage of students in Proficiency Levels by state and territory in 2012

State/ Territory	Level 2 and below	Level 3.1	Level 3.2*	Level 3.3	Level 4 and above	At or above the Proficient Standard
ACT	4.4 (±1.7)	30.3 (±4.6)	49.4 (±4.2)	15.0 (±4.1)	0.9 (±0.7)	65.3 (±5.3)
NSW	9.2 (±2.5)	39.8 (±3.3)	40.9 (±3.8)	9.6 (±2.5)	0.4 (±0.4)	50.9 (±4.3)
NT	31.1 (±9.6)	37.9 (±7.0)	26.5 (±6.4)	4.3 (±3.0)	0.2 (±0.4)	31.0 (±7.6)
QLD	8.8 (±1.6)	41.4 (±2.9)	41.8 (±3.1)	8.0 (±1.6)	0.1 (±0.2)	49.9 (±3.3)
SA	8.8 (±1.9)	40.1 (±3.4)	43.5 (±3.6)	7.5 (±1.9)	0.1 (±0.2)	51.1 (±3.9)
TAS	9.6 (±2.3)	39.1 (±4.2)	40.2 (±4.2)	10.8 (±3.2)	0.3 (±0.5)	51.3 (±5.4)
VIC	8.3 (±2.2)	40.4 (±4.0)	43.4 (±3.8)	7.6 (±2.3)	0.2 (±0.3)	51.3 (±4.7)
WA	8.2 (±1.9)	35.5 (±3.3)	44.0 (±3.3)	12.0 (±2.4)	0.4 (±0.4)	56.4 (±4.2)
AUST	9.0 (±1.0)	39.6 (±1.6)	42.1 (±1.7)	9.0 (±1.1)	0.3 (±0.2)	51.4 (±2.0)

Notes: *The Proficient Standard has been set at Proficiency Level 3.2.
Figures in parentheses refer to 95 per cent confidence intervals.

Appendix Four

Comparison of Males and Females Raw and scaled scores across Australian States.

Note: ( ) bracketed numbers indicate the sample size.

STATE or Territory	National results Male		National Results Female		IB Results Male		IB Results Female	
	Raw score/51	Scaled score	Raw score/51	Scaled score	Raw score/51	Scaled score	Raw score/51	Scaled score
ACT	26.4	432	25.6	425	32.45 (20)	504	29.3 (22)	465
NSW	23	395	22.9	394	28.66 (73)	456	No results	No results
QLD	22.6	391	22.7	392	27.6 (8)	447	29.08 (12)	463
SA	22.6	391	22.8	393	27.5 (49)	445	24.5 (12)	411
VIC	22.7	392	23	395	29.3 (29)	465	27.45 (31)	444
WA	24	405	24.1	407	31.27 (55)	489	No results	No results
Overall	23	394	23	395	29	462	28	451