RESEARCH SUMMARY

PYP and MYP Student Performance on the International Schools' Assessment (ISA)
Australian Council for Educational Research

Summary prepared by IB Global Policy & Research Team
March 2010

ABSTRACT

This study, undertaken by the Australian Council for Educational Research (ACER), investigated how International Baccalaureate (IB) students enrolled in the Primary Years Programme (PYP) and Middle Years Programme (MYP) performed on the International Schools' Assessment (ISA), relative to non-IB students. The ISA assesses student performance in Grades 3 to 10 across four domains: *Math Literacy, Reading, Narrative Writing, and Expository Writing*. The math and reading components of the assessment are based on the reading and mathematical literacy frameworks of the OECD's Programme for International Student Assessment (PISA). The study sample included IB students (*N*=23,575) and non-IB students (*N*=14,317) across Asia and Oceania, Europe, Africa, and the Americas, who participated in the ISA in 2007/2008 and 2008/2009. On the whole, despite some regional differences, the results indicate that IB PYP and MYP students outperformed their non-IB peers on the ISA across all four domains in a majority of grade levels, with the strongest effects noted in Year 10 Math and Expository Writing. IB students' ISA scores in Grades 9 and 10 also compare very favourably to PISA benchmarks in Math and Reading. On the other hand, there was insufficient evidence to suggest that IB schools authorized for a longer period produce better student outcomes, and no clear patterns were noted in student performance across IB full continuum schools and single or dual programme schools.

PROJECT OVERVIEW

In 2009, the IB contracted the Australian Research Council for Educational Research (ACER) to undertake an analysis of the International Schools' Assessment (ISA) data for participating IB schools.

The ISA is an assessment developed for international school students in Grades 3 to 10. The ISA uses a combination of multiple choice and open-ended questions, across four domains: *Math Literacy*, *Reading*, *Writing Task A* (narrative writing), and *Writing Task B (expository writing)*. The math and reading portions are based on the OECD's Programme for International Student Assessment (PISA) math and reading literacy frameworks. The ISA is administered twice a year, in October and the following February. Approximately 60-70% of the schools participating in the ISA administration implement the IB curriculum, thus providing an opportunity to examine the performance of PYP and MYP students in Grades 3 to 10, relative to non-IB students. The analysis pertinent to this study focuses on students' performance in four sittings of the ISA conducted in 2007/2008 and 2008/2009.

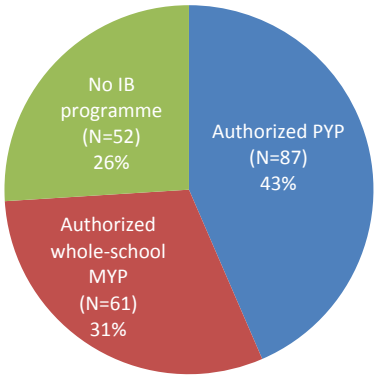
The analyses undertaken in this study were designed to address the following research questions:

1) How do IB PYP and MYP students perform on the four ISA domains, relative to a non-IB comparison group?
2) How do IB students' ISA scores in grade 9 and grade 10 compare with PISA benchmarks?
3) How do IB MYP moderation results align with ISA scores?
4) What is the relationship, if any, between length of authorization and student performance on the ISA?
5) How do IB PYP and MYP students in full continuum schools[1] perform on the ISA, as compared to students in schools with one or two IB programmes?

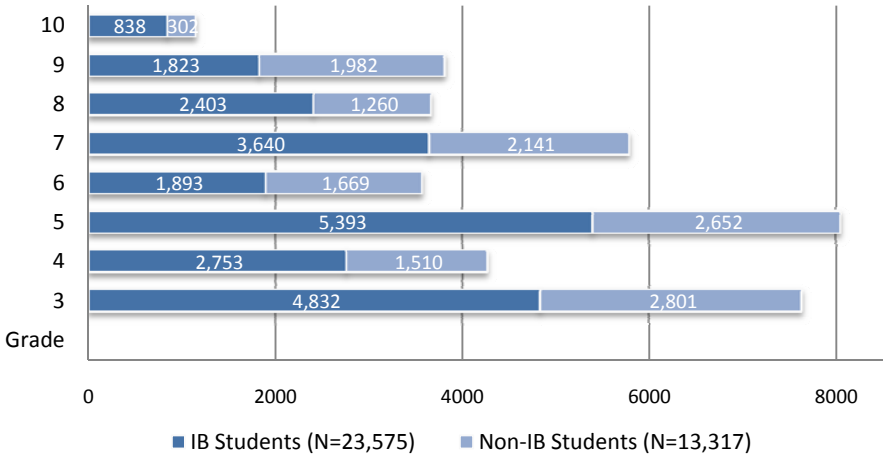
In the sections that follow, we provide a demographic overview of the study sample, followed by key findings pertinent to the aforementioned research questions.

ISA STUDY SAMPLE

Participant schools, by programme

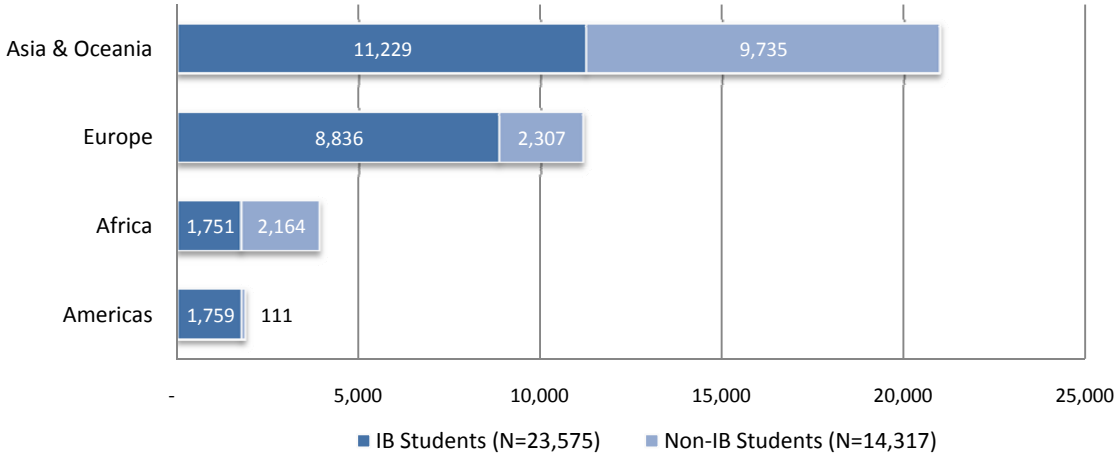


Participant students, by Grade



[1] The term "full continuum schools" refers to schools that offer all three IB programmes.

Participant Students, by Region



For the purposes of this study, an IB school refers to those with authorized PYP implementation in Grades 3 to 5, and authorized whole-school MYP implementation in Grades 6 to 10. An IB student is a student enrolled in an IB school. A non-IB school refers to those with no IB programme, and excludes 'IB-candidate', 'IB-interested' and 'IB-withdrawn' schools. A non-IB student is a student enrolled in a non-IB school.

KEY FINDINGS

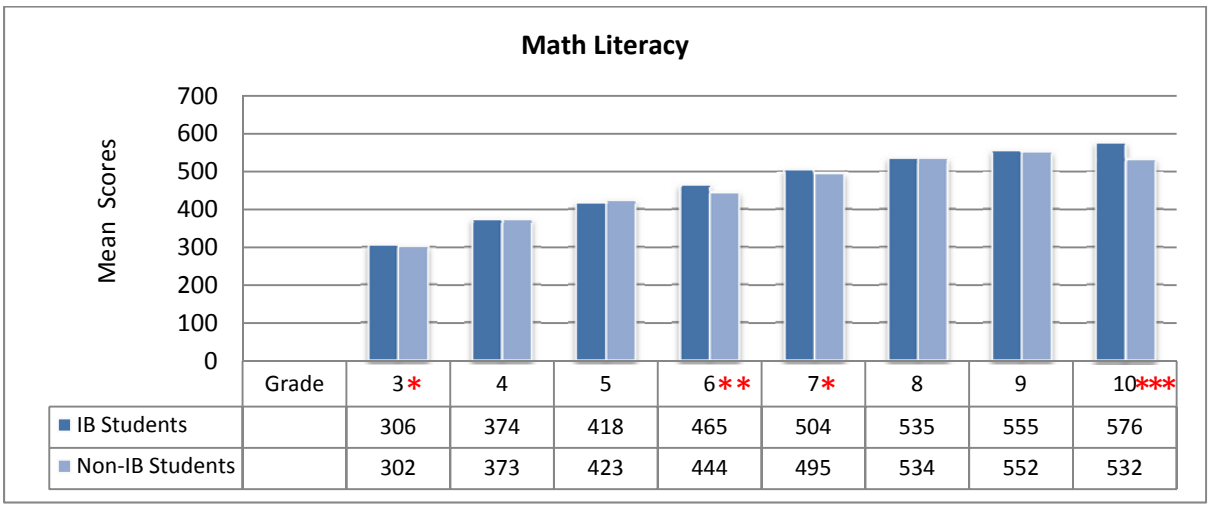
1. How do IB PYP and MYP students perform on the ISA, relative to a non-IB comparison group?

1.1 Analysis of student performance across four ISA domains

At a broad level, results indicate that **IB PYP and MYP students performed as well as or better on the ISA than their non-IB peers across all four domains in a majority of grade levels**, as illustrated below in Figures 1 to 4.

As evident from Figure 1, IB students outperformed their non-IB peers to a large extent in Grade 10 Math Literacy. In terms of Math sub-strands, IB students in Grade 10 performed significantly better than their non-IB peers in (i) *Change and relationships*, (ii) *Quantity*, and (iii) *Uncertainty*. ISA sub-strand descriptions are provided in Appendix A.

Figure 1: IB and non-IB Student Performance on Math Literacy



Grade	3*	4	5	6**	7*	8	9	10***
IB Students	306	374	418	465	504	535	555	576
Non-IB Students	302	373	423	444	495	534	552	532

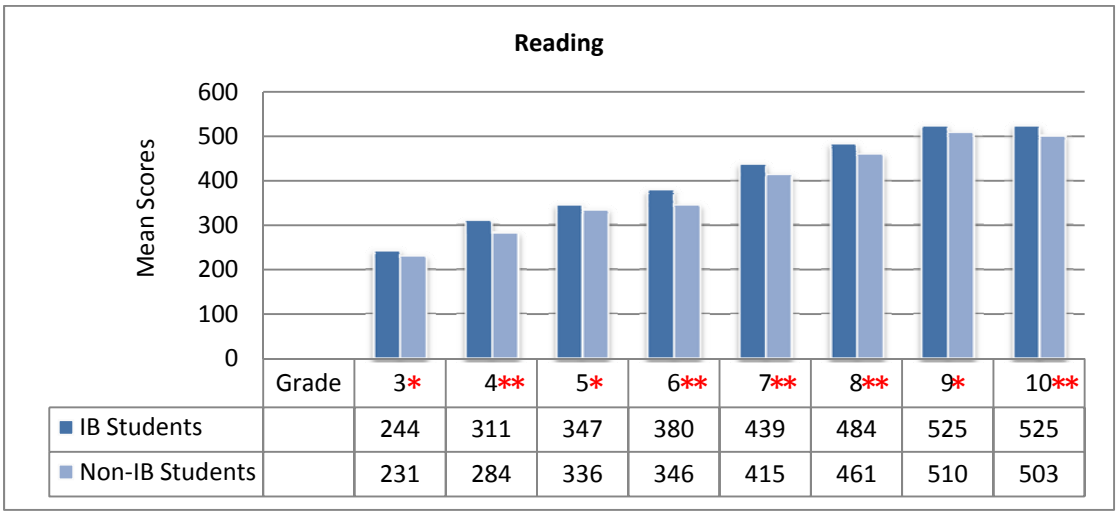
* Statistically significant difference (higher), **small** effect size ($0.1 \leq d < 0.2$)

** Statistically significant difference (higher), **moderate** effect size ($0.2 \leq d < 0.5$)

*** Statistically significant difference (higher), **large** effect size ($d \geq 0.5$)

As shown in Figure 2, IB students performed consistently better than their non-IB peers in Reading across all grade levels, with the differences being most significant (moderate effect size) at Grades 4, 6, 7, 8, and 10. In terms of Reading sub-strands, IB students performed particularly well in *Reflecting* (Grade 4 and 10) and *Retrieving Information* (Grade 10).

Figure 2: IB and non-IB Student Performance on Reading

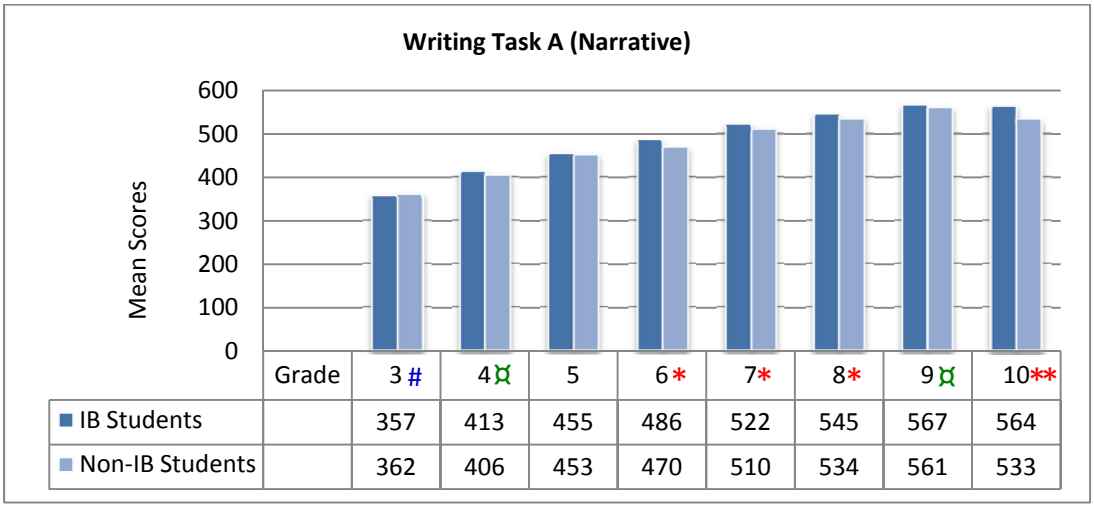


Grade	3*	4**	5*	6**	7**	8**	9*	10**
IB Students	244	311	347	380	439	484	525	525
Non-IB Students	231	284	336	346	415	461	510	503

* Statistically significant difference (higher), **small** effect size ($0.1 \leq d < 0.2$)

** Statistically significant difference (higher), **moderate** effect size ($0.2 \leq d < 0.5$)

Figure 3: IB and non-IB Student Performance on Writing Task A (Narrative)



Grade	3 #	4 ¤	5	6 *	7 *	8 *	9 ¤	10 **
IB Students	357	413	455	486	522	545	567	564
Non-IB Students	362	406	453	470	510	534	561	533

\# Statistically significant difference (lower), **negligible** effect size ($d < 0.1$)

¤ Statistically significant difference (higher), **negligible** effect size ($d < 0.1$)

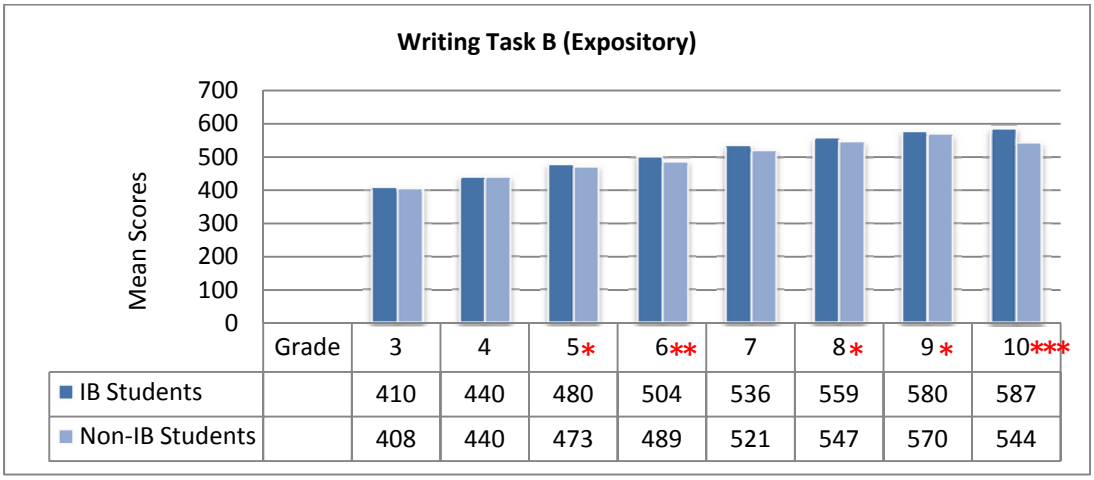
* Statistically significant difference (higher), **small** effect size ($0.1 \leq d < 0.2$)

** Statistically significant difference (higher), **moderate** effect size ($0.2 \leq d < 0.5$)

As illustrated in Figure 3, IB students' performance on Writing Task A (Narrative) was found to be less uniform than the comparative results found for Math Literacy or Reading. While IB students generally performed as well as, if not better, than their non-IB peers across most grade levels, they scored a marginally lower mean score in Grade 3. It is important to note, however, that the effect size (d=-0.07) suggests that this difference is negligible. In terms of Narrative Writing sub-strands, IB students outperformed non-IB students considerably in *Spelling* (d=0.57) and *Content* (d=0.39), particularly in Grade 10.

On Writing Task B (Expository), Figure 4 shows that IB students performed as well as, if not significantly better, than their non-IB peers across all grade levels, with the largest differences noted in Grade 10. In terms of Expository Writing sub-strands, IB students outperformed non-IB students to a moderate degree in (i) *Content*, (ii) *Language*, and (iii) *Structure and Organization*.

Figure 4: IB and non-IB Student Performance on Writing Task B (Expository)



Grade	3	4	5*	6**	7	8*	9*	10***
IB Students	410	440	480	504	536	559	580	587
Non-IB Students	408	440	473	489	521	547	570	544

* Statistically significant difference (higher), **small** effect size ($0.1 \leq d < 0.2$)

** Statistically significant difference (higher), **moderate** effect size ($0.2 \leq d < 0.5$)

*** Statistically significant difference (higher), **large** effect size ($d \geq 0.5$)

1.2 Regional analysis of student performance across four ISA domains

For the purposes of the regional analysis, due to the small number of schools and students in Americas and Oceania, the former is grouped with Europe, and the latter with Asia. The ISA student performance comparisons are thus presented in three geographical regions, namely (i) Europe and Americas, (ii) Africa, and (iii) Asia and Oceania.

On the whole, **IB students in (i) *Africa,* and (ii) *Europe & Americas* performed as well as, if not better than, their non-IB peers across all four domains and in most grades**, with the largest effect sizes in Reading and Math Literacy.

On the other hand, the ***Asia & Oceania* results demonstrate a more varied pattern in IB students' performance relative to a non-IB comparison group**. In this region, IB students performed as well as, and occasionally better than, their non-IB peers across three domains: Reading, Writing Tasks A (Narrative) and B (Expository). In terms of Math Literacy, however, IB students demonstrated stronger performance in Grade 10, but weaker performance in Grades 4, 5, 8, and 9. These results are illustrated in Figures 5 to 7.

Figure 5: IB and non-IB Student Performance in Europe & Americas

Note: A total of 32 comparisons were made (i.e. Grades 3-10 x 4 domains). IB students performed stronger, or as well as their non-IB peers in 24 instances and 8 instances, respectively.

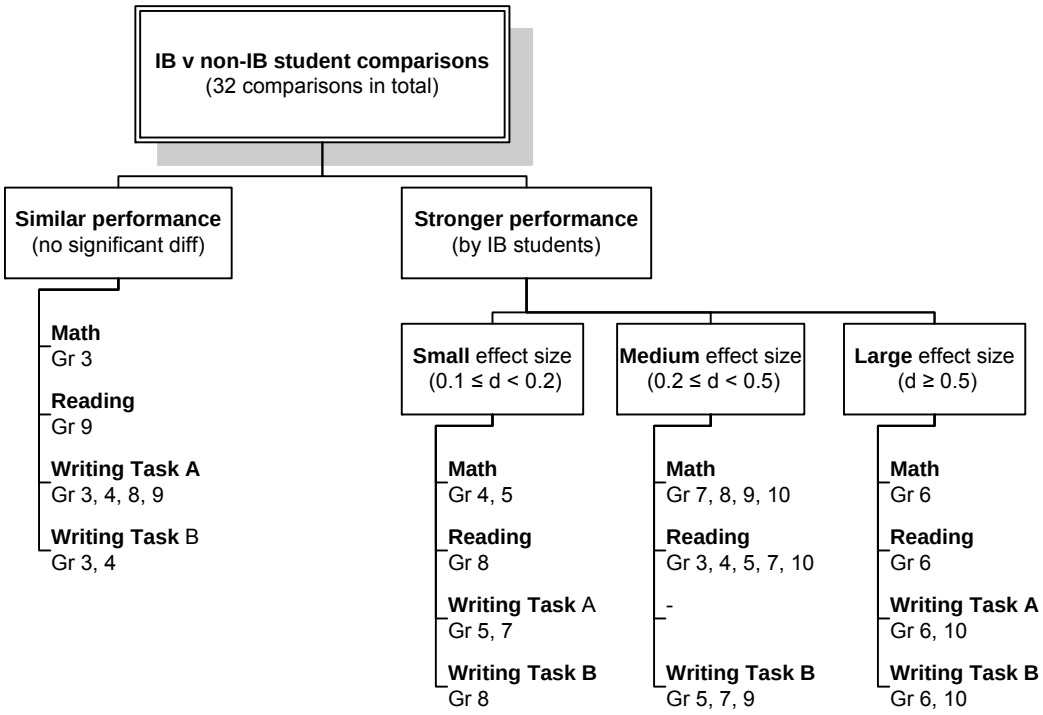
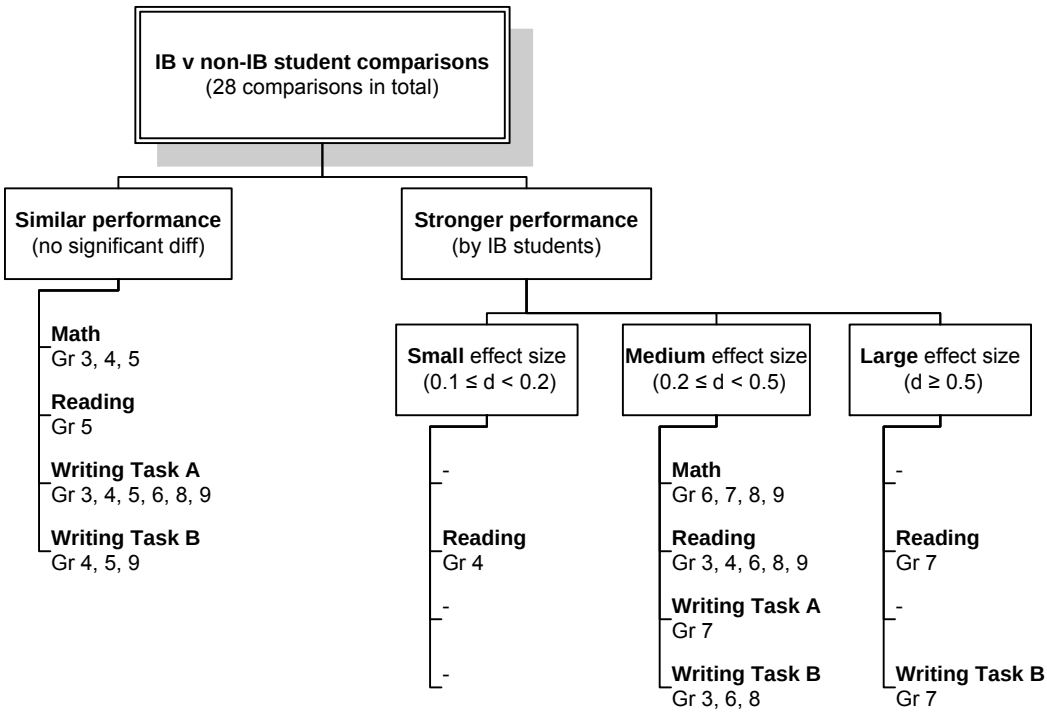


Figure 6: IB and non-IB Student Performance in Africa[2]

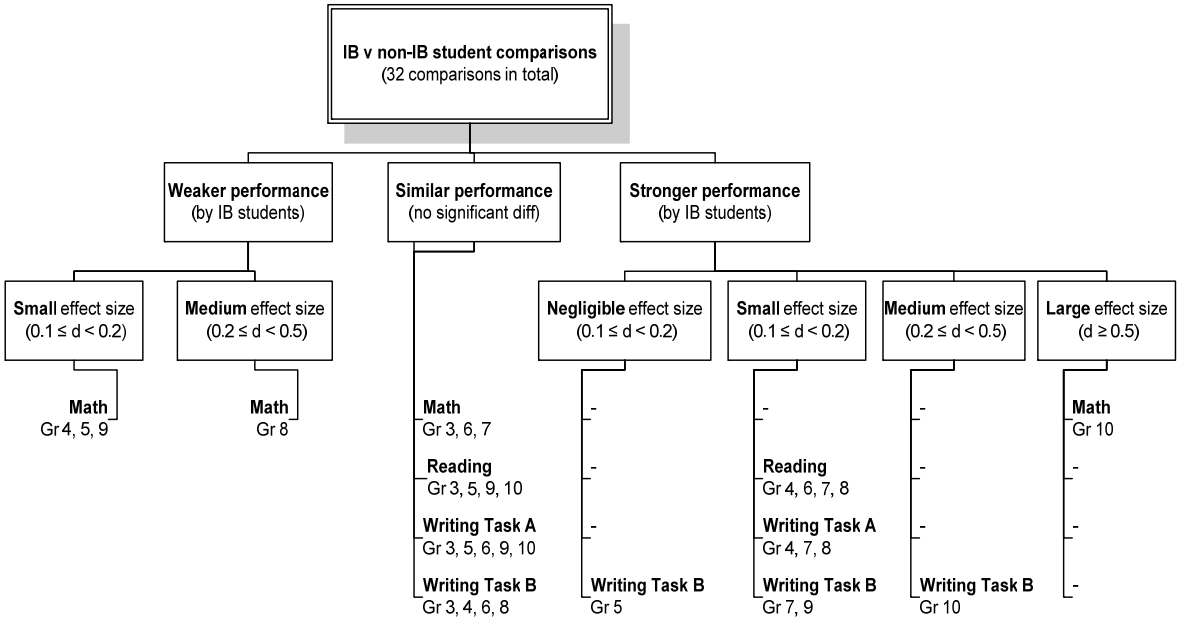
Note: A total of 28 comparisons were made (i.e. Grades 3-9 x 4 domains). IB students performed stronger, or as well as their non-IB peers in 15 instances and 13 instances, respectively.



[2] No comparisons at Grade 10 level were made because there are no IB students at Grade 10 in Africa.

Figure 7: IB and non-IB Student Performance in Asia & Oceania

Note: A total of 32 comparisons were made (i.e. Grades 3-10 x 4 domains). IB students performed stronger, or as well as their non-IB peers in 12 instances and 16 instances, respectively. IB students demonstrated weaker performance as compared to their non-IB peers in 4 instances.



2. How do IB students' ISA scores compare with PISA benchmarks?

To address this research question, ISA scores of Grade 9 and 10 IB students in the domains of Math Literacy and Reading were compared with the 2006 PISA mean scores of OECD countries and partner countries.

As shown in Figure 8, for **Math Literacy, IB students in both Grade 9 and 10 performed significantly better than the OECD mean** (with large effect sizes in the range of 0.53 to 1.01), and **better than all countries that participated in the PISA Mathematics**. Specifically, the average 2006 OCED score ranged from 447 to 547, while the IB Grade 9 and 10 average scores were 555 and 576 respectively.

As shown in Figure 9, for **Reading, IB students in both Grade 9 and 10 performed better than most of the PISA participant countries**, with the exception of Finland, Korea, Hong Kong/China. The average OCED score ranged from 447 to 556, while the IB average scores were 525 for both Grade 9 and 10.

Figure 8: Grade 9 and 10 IB Students' Performance in Math Literacy, relative to PISA benchmark[3]

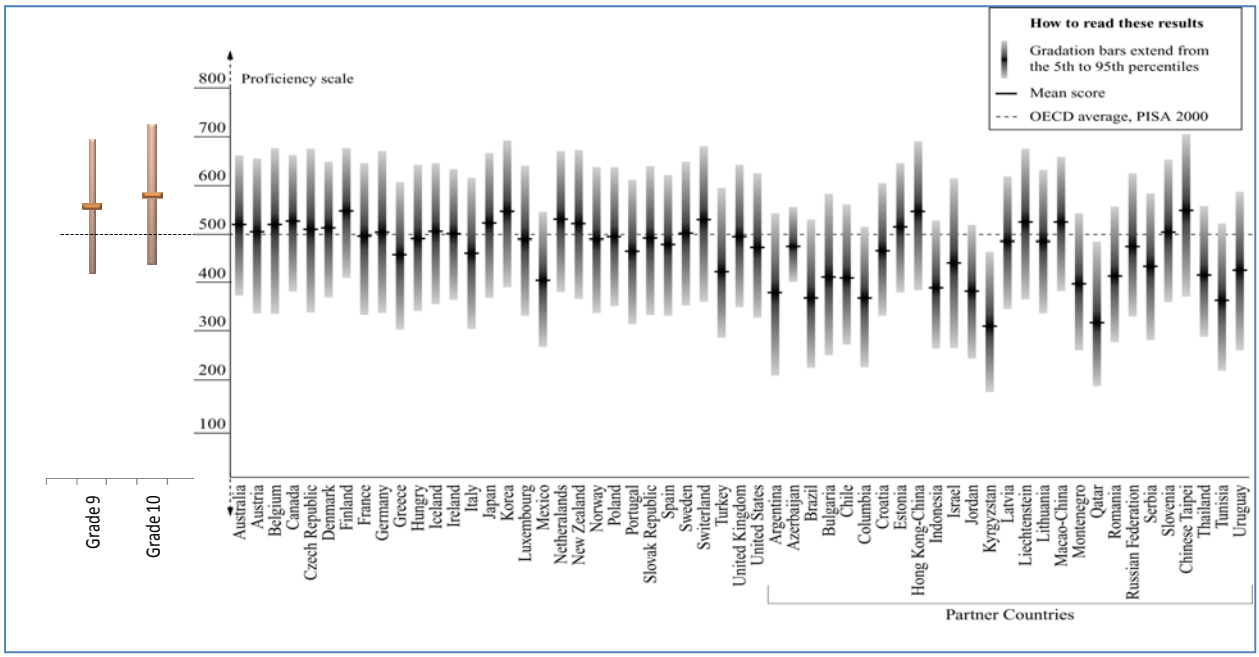
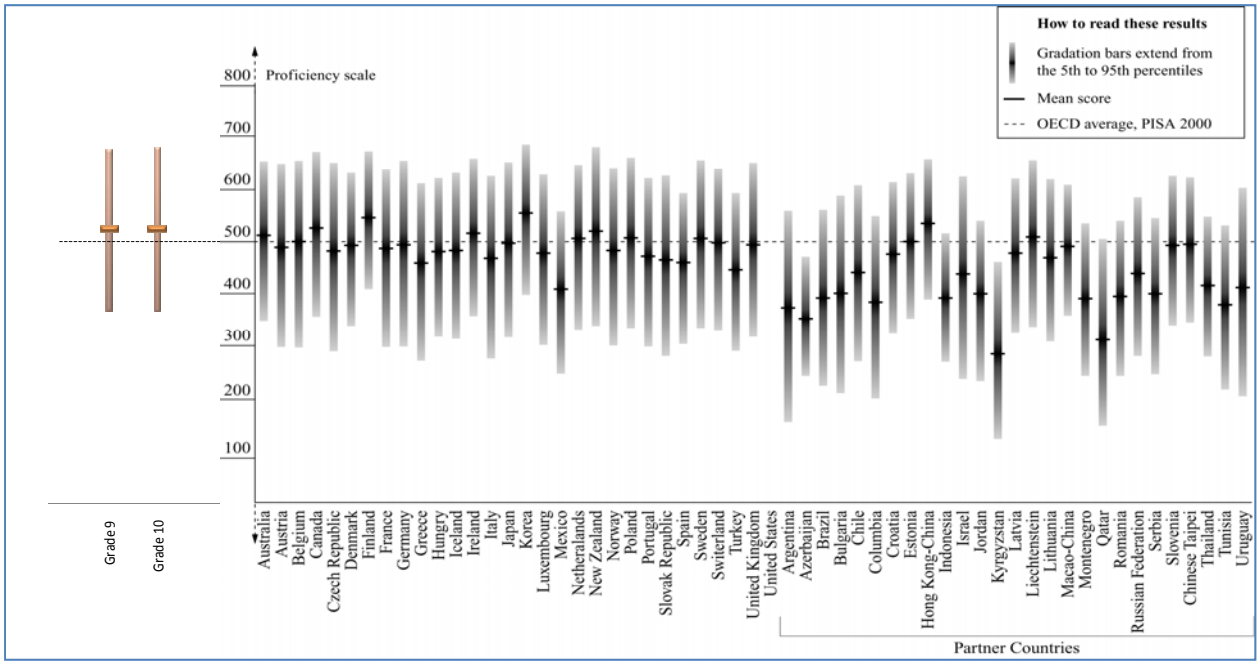


Figure 9: Grade 9 and 10 IB Students' Performance in Reading, relative to PISA benchmark[3]



[3] The bars depicted in Figures 8 and 9 refer to IB students' Grade 9 and 10 ISA scores and 2006 OECD average scores. The dotted line depicting the 2000 OECD average score as 500 is used as a reference point to facilitate ease of interpretation.

3. How do IB MYP moderation results align with ISA scores?

In general, the results indicate that **there are common skills underlying IB MYP moderation and three ISA assessment areas**, namely, Math Literacy, Reading, and Writing Task B (Expository).

To evaluate how MYP moderation performance was aligned with ISA scores, two forms of regression analyses were conducted to examine (i) concurrent correlation, and (ii) predictive correlation.

The concurrent correlation analysis focused on whether the ISA assessment scores in Grade 10 could explain the MYP moderation score measured at approximately the same time. This regression analysis (see Figure 10: '2008 Grade 10') was conducted using the *average MYP moderation score*[4] as a dependent variable, and the *ISA assessment scores*[5] as independent variables.

The predictive correlation analysis examined the extent to which ISA assessment scores in Grade 9 predicted the average MYP moderation score achieved a year later. This regression analysis (see Figure 10: '2007 Grade 9') was conducted using the *Grade 9 ISA assessment scores* as predictor variables and the *average MYP moderation score* in Grade 10 as the outcome variable.

Figure 10: Regression model summary[6] results of Average MYP Moderation Score and ISA Scores

Regression models	R	R Square	Adjusted R Square	Std. Error of the Estimate
2007 Grade 9	0.636	0.404	0.397	0.64463
2008 Grade 10	0.624	0.39	0.378	0.57775

a. Dependent Variable: avg_MYP , and predictors: (Constant), Write Task B, grade, Mathematical Literacy, Reading

The adjusted R^2 results show that approximately 40% of the variance in the average MYP moderation scores can be explained by the regression model with predictors being ISA scores in Math Literacy, Reading, and Writing Task B. **This indicates that there are common skills underlying MYP moderation and these three ISA assessment areas.**

4. What is the relationship between length of authorization and student performance on the ISA?

In sum, there was **insufficient evidence to suggest that schools authorized for a longer period of time produced better student performance outcomes**.

To assess the relationship between length of authorization and student performance on the ISA, correlation analyses were conducted between the *length of IB implementation since authorization to 1st March 2009* and the *ISA mean scores* across all four assessment domains: Math Literacy, Reading, Writing Tasks A and B.

As illustrated in Table 1, only a handful of statistically significant correlations[7] (weak to moderate) were noted in the assessment domains of Mathematical Literacy, Reading, and Writing Task A. Specifically,

[4] Calculated as a mean score of at least 5 MYP moderation subjects.
[5] All four ISA assessment scores were included in the initial analyses, but Writing Task A was dropped from the list of independent variables subsequent to the preliminary analysis which established its coefficient as statistically insignificant.
[6] Descriptive statistics, correlations, coefficients and residual statistics are not reported here but can be made available on request.
[7] Significant at the 95% confidence level.

weak to moderate correlations were noted between length of PYP implementation and Grade 5 Math (0.25, p<0.05), Grade 4 Reading (0.28, p=0.05) and Grade 5 Reading (0.36, p<0.05). Also weak to moderate correlations were noted between length of MYP implementation and Grade 9 Reading (0.50, p<0.00) and Grade 8 Writing Task A (0.34, p=0.05). These statistically significant correlations are bolded/highlighted in Table 1 below. No other statistically significant correlations were found.

Table 1: Correlation analyses between length of IB implementation and ISA scores

Grade	Domain	IB Implementation Year			Domain	IB Implementation Year		
		r	p<	n		r	p<	n
3	Mathematical Literacy	0.12	0.30	72	Writing Task A	-0.00	0.89	72
4		0.19	0.19	49		0.03	0.82	49
5		**0.25**	**0.03**	77		0.19	0.10	76
6		0.24	0.21	29		0.18	0.36	29
7		0.11	0.47	49		0.03	0.84	49
8		0.31	0.07	34		**0.34**	**0.05**	34
9		**0.36**	**0.05**	32		0.30	0.09	32
10		0.11	0.77	10		0.44	0.20	10
3	Reading	0.19	0.11	72	Writing Task B	0.05	0.65	72
4		**0.28**	**0.05**	49		0.25	0.08	49
5		**0.30**	**0.01**	77		0.22	0.06	77
6		0.18	0.36	29		0.11	0.57	29
7		0.07	0.63	49		0.03	0.81	49
8		0.32	0.07	34		0.27	0.12	34
9		**0.50**	**0.00**	32		0.30	0.09	32
10		0.09	0.82	10		0.30	0.40	10

5. *How do IB PYP and MYP students in full continuum schools perform on the ISA, as compared to students in schools with one or two IB programmes?*

On the whole, **no clear patterns were identified in the performance of IB full continuum schools as compared to schools with one or two programs**. Specifically, the following comparisons were made:

1) How do students' ISA performance compare between IB full continuum schools and non-IB schools?
2) How do students' ISA performance compare between IB full continuum schools and single-program schools (i.e., PYP-only and MYP-only)?
3) How do students' ISA performance compare between IB full continuum schools and dual-program schools (i.e., PYP+MYP, MYP+DP, and PYP+DP)?

Findings pertaining to each of these comparisons are described in the following sub-sections.

5.1 IB full continuum schools and non-IB schools

Students from IB full continuum schools outperformed non-IB school students across all domains in Grades 6 to 10. In particular, large effect sizes were noted in Grade 10 Math Literacy (*d*=0.61) and Writing Task B (*d*=0.57).

In all other grades, **students from IB full continuum schools performed as well as non-IB students across all domains**. Only one exception was noted in the case of Grade 3 Writing Task A, where IB students performed marginally weaker (*d*=0.12) than non-IB students.

5.2 IB full continuum schools and single-program schools

Generally, **no significant differences were noted between the ISA performance of students enrolled in IB full continuum and single-program PYP or MYP schools**. Specifically, of 24 comparisons[8], only 3 statistically significant differences were noted, where students from IB full continuum schools performed weaker than their IB peers from single-program PYP or MYP schools. These were:

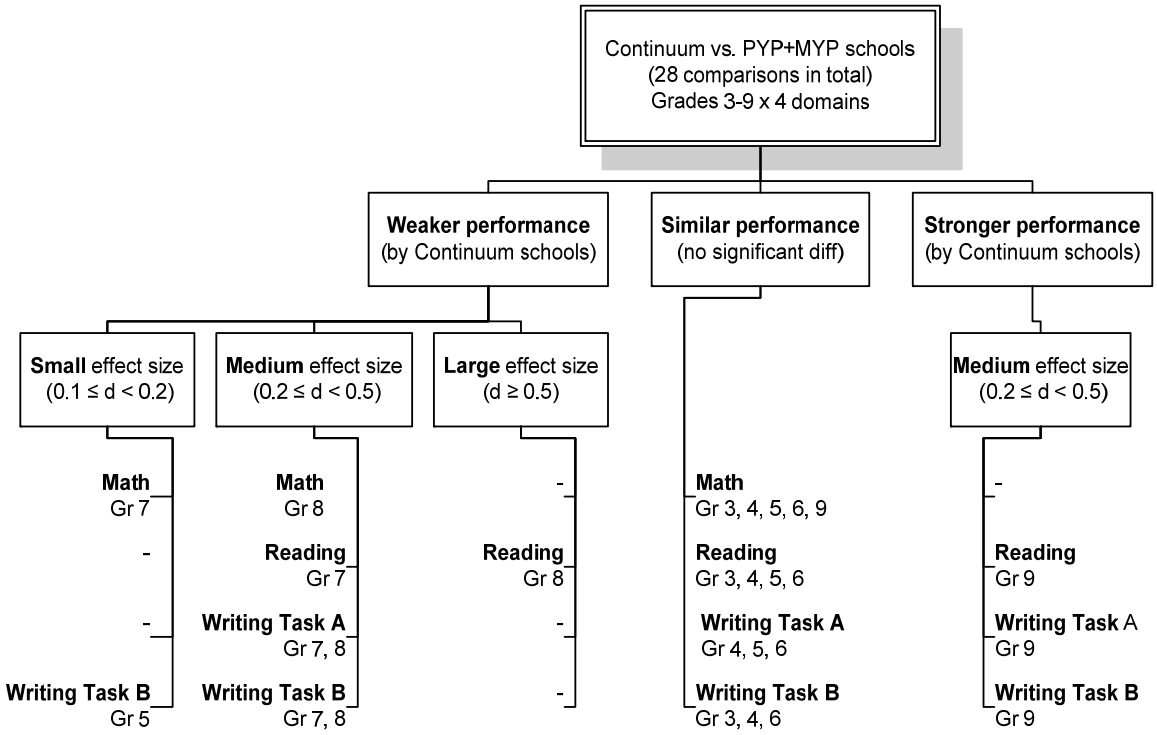
- Grade 3 Reading (medium effect size, *d*=0.26)
- Grade 5 Writing Task A (small effect size, *d*=0.12)
- Grade 10 Writing Task B (large effect size, *d*=0.65)

In all other cases, students from single-program IB schools achieved marginally higher mean scores than their peers from IB full continuum schools, but the differences were not statistically significant.

5.3 IB full continuum schools and dual-program schools

No clear patterns were found in the ISA performance of students from IB full continuum schools and dual-program schools. In different domains across different grades, students from IB full continuum schools performed similarly, weaker, or stronger than their peers from dual-program schools. Results specific to the comparisons with PYP+MYP schools, MYP+DP schools and PYP+DP schools are shown in Figures 11 to 13 on the following pages.

Figure 11: Comparison of IB full continuum schools and PYP+MYP schools[9]



[8] Student performance data in Grades 3-5, 7, 9 10 across four ISA domains were included in this analysis, thereby giving a total of 24 comparisons. ISA assessment data for schools with only an MYP in Grade 6 were not available, and therefore Grade 6 was excluded from the analysis. Grade 8 was also excluded from the analysis due to an insufficient sample (ie, only 10 students).

[9] No comparisons were made at the Grade 10 level because no ISA assessment data was available for schools with the two programmes (PYP+MYP) in Grade 10.

Figure 12: Comparison of IB full continuum schools and MYP+DP schools[10]

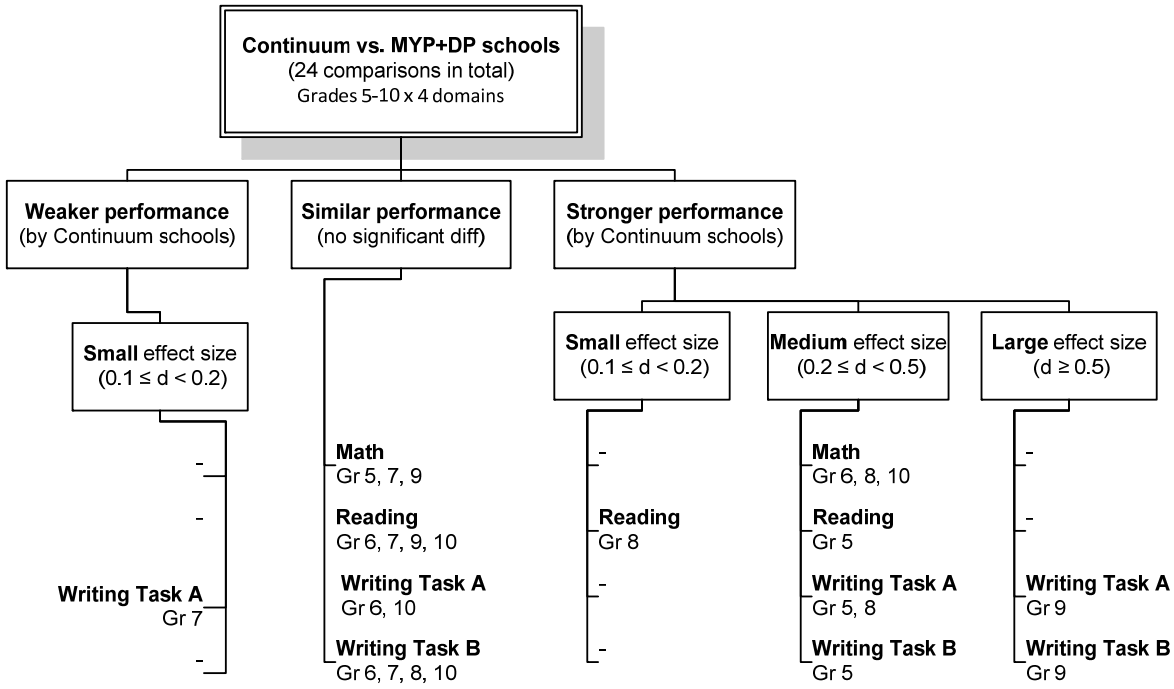
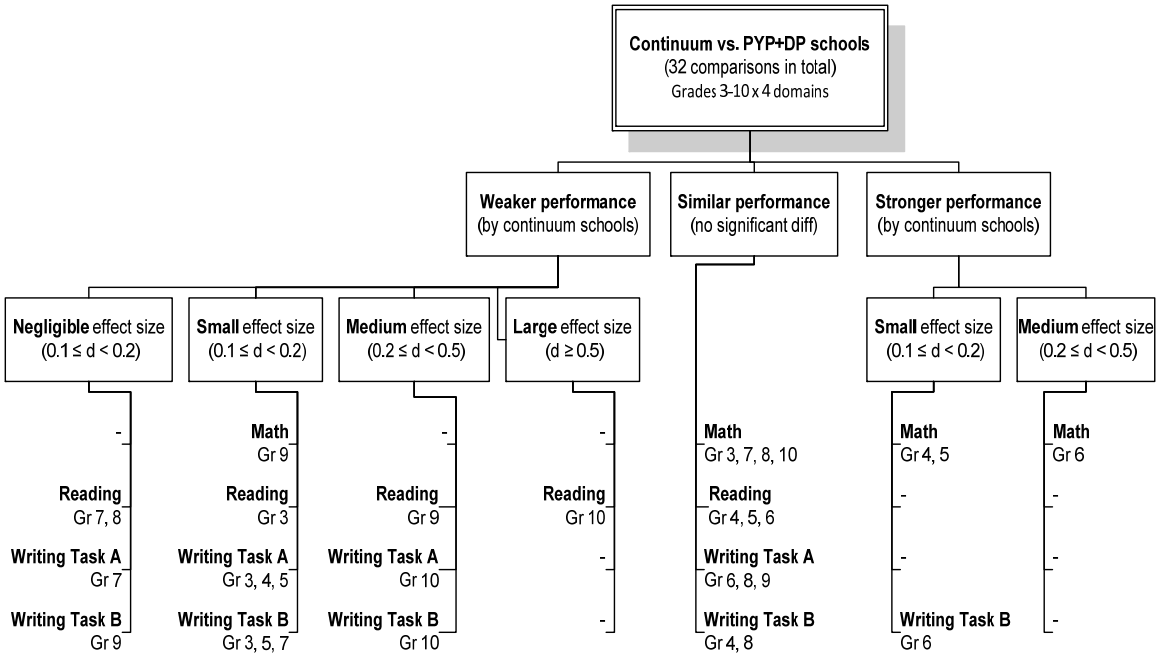


Figure 13: Comparison of IB full continuum schools and PYP+DP schools



[10] No comparisons were made at the Grade 3 and 4 levels, because ISA participant schools with MYP+DP programmes exist only in Grades 5-10.

CONCLUSION

To date, this is the first international study examining how IB students performed on the ISA as compared to their non-IB peers across Asia, Oceania, Europe, Americas, and Africa.

The results that emerged from this study are generally positive and encouraging. On the whole, despite some regional differences in the results, IB students tended to outperform their non-IB peers on the ISA across all domains in a majority of grade levels. In addition, IB students' ISA scores compare favourably to PISA benchmarks. In other words, IB students' Grade 9 and 10 ISA scores in Math and Reading are significantly higher than most, if not all, OECD participant and partner countries' mean scores in the same assessment areas. With regards to the alignment of MYP moderation results and ISA scores, the results suggest common skills underlying the assessment areas of Math, Reading, and Writing Task B (expository). On the other hand, there was insufficient evidence to suggest that schools authorised for a longer period of time produce better student outcomes, and no clear patterns were noted in student performance across IB full continuum schools and single or dual program schools.

Future research efforts will aim to replicate this study over multiple ISA assessment periods to obtain a longitudinal analytical trend, as well as to explore further some of the less intuitive findings around length of authorisation and IB full continuum schools. In addition, the IB research team is currently exploring the possibilities of engaging ACER to administer a learner questionnaire measuring non-academic educational outcomes in conjunction with the ISA administration in October 2010 and February 2011. The purpose of this extension is to move beyond student performance to gain empirical insights into IB students' perceptions of school life, values and attitudes, as well as social and emotional well-being, among others.

This summary was developed by the IB's Global Policy & Research Department. For more information on this study, or other IB research, e-mail the IB Research Department at research@ibo.org.

To cite the full report, please use the following:

Tan, L. & Bibby, Y. (2010). IB PYP and MYP student performance on the International Schools' Assessment *(ISA)*. Melbourne: Australian Council for Educational Research.

APPENDIX A: ISA Sub-strand Description

Detailed descriptions of ISA sub-strands in Mathematical Literacy, Reading, and Writing Tasks A and B provided below have been extracted from ACER's *Guide to ISA Reports for October 2008 Administration*.

Mathematical Literacy Sub-strands

Uncertainty This content area reflects how in real life data is commonly collected, organized, analysed and displayed with a view to making interpretations and forming conclusions. Many decisions are made based upon statistical analysis of data. Real life also contains elements of chance where outcomes are not certain but based upon probabilities. Increasingly decision-making is qualified with a statement of risk and society is presented with more and more information to make sense of.

Quantity This overarching content area also features in the three other domains to varying degrees. It focuses on the need for quantification in order to organize the world. It is not hard to find examples of quantification in our day-to-day living. We use money, make measurements, estimate, and calculate. Increasingly we make use of technology to assist us but we also still perform many calculations mentally and approximately. Quantitative reasoning requires number sense: that is, having a feel for the magnitude of numbers, using strategies and tools appropriately, and being able to check solutions for reasonableness.

Space and Shape Shapes and constructions are all around us physically as real objects but also as representations in the form of photographs, maps, and diagrams. Constructing and interpreting such representations is an important skill. Using known geometric shapes whose mathematical properties are known to model more complex shapes is an important problem-solving tool. Knowledge and appreciation of the beauty and function of geometric shapes and spaces has applications reaching from art to advertising.

Change and Relationships Noticing and using patterns in number and shapes, and finding and describing relationships between variables lies at the heart of mathematics. As organisms or populations grow and as stock markets ebb and flow, we describe the patterns in words, in tables and sometimes in algebraic notation. Commonly we chart the changes in graphical form. These descriptions can be linear, non-linear, cyclic, and exponential to name but a few. Being able to link between these various representations and use the language, notation, and algorithms of change and relationships is critical to making sense of the patterns in our world.

Reading Sub-strands

Retrieving Information is defined as locating one or more pieces of information in a text.

Interpreting texts is defined as constructing meaning and drawing inferences from one or more parts of a text.

Reflecting is defined as relating a text to one's experience, knowledge, and ideas.

Criteria for Writing Task A (Narrative)

Narrative – Content criterion is about the quality and range of ideas presented, the development of plot, characters and setting, and the writer's sense of audience and purpose. It also encompasses the overall shaping of the piece.

Narrative – Language criterion deals with sentence and paragraph structure, vocabulary, and punctuation, and the writer's voice.

Narrative – Spelling criterion takes into account students' knowledge of phonetic and visual spelling patterns and the range of words attempted, as well as correctness of spelling.

Criteria for Writing Task B (Exposition or Argument)

Exposition/Argument – Content criterion looks at the depth and range of ideas presented, and at the quality of reasoning demonstrated in the ability to provide evidence and logical argumentation in support of a position.

Exposition/Argument – ESOL Language (English for Speakers of Other Languages) criterion is applied to all students' writing regardless of their language background, but focuses on the grammatical correctness and command of English syntax, as well as sentence fluency and variation, and vocabulary.

Exposition/Argument – Structure and Organisation criterion deals with the overall structure of the writing, for example the presence of a clear introduction, development and conclusion; and its internal coherence, such as linking between and within paragraphs.