ACER

Performance comparison between IB and Non-IB students on the International Schools' Assessment: Applying hierarchical linear modelling to the 2023 and 2024 data

Liang-Cheng Zhang

Report prepared for the International Baccalaureate by ACER 2026

Performance comparison between IB and Non-IB students on the International Schools' Assessment: Applying hierarchical linear modelling to the 2023 and 2024 data

Liang-Cheng Zhang

Australian Council for Educational Research Limited
19 Prospect Hill Road
Camberwell, Victoria 3124
Australia
ABN 19 004 398 145

www.acer.org

Recommended APA 7 citation

Zhang, L.-C. (2026). *Performance comparison between IB and Non-IB students on the International Schools' Assessment: Applying Hierarchical Linear Modelling to the 2023 and 2024 data.* International Baccalaureate Organization. https://ibo.org/research/outcomes-research/pyp-studies/performance-comparison-between-ib-and-non-ib-students-on-the-international-schools-assessment-2026

Table of Contents

Table of Contents ii

List of Figures iv

List of Tables v

Abbreviations and Acronyms vi

Executive Summary 1

1. Project Overview 3

1.1 Information on IB Schools and Students Participating in the ISA 4

1.2 Methodology 5

1.2.1 Hierarchical (multilevel) Linear Models 5

1.2.2 Effect Size 7

2. Analysis of Student Performance 9

2.1 How do PYP and MYP students compare to non-IB students from schools accredited for international education on the ISA assessment areas? 9

2.1.1 Mathematical Literacy 9

2.1.2 Reading 10

2.1.3 Scientific Literacy 11

2.1.4 Narrative Writing 12

2.1.5 Expository Writing 13

2.2 What is the magnitude of the effect? 14

2.2.1 Mathematical Literacy 15

2.2.2 Reading 16

2.2.3 Scientific Literacy 17

2.2.4 Narrative Writing 18

2.2.5 Expository Writing 19

2.3 How do the Grades 9 and 10 ISA scores of IB students align with PISA benchmarks? 21

3. Conclusion 23

4. Appendix 24

4.1 Density Plots of Non-Modelled Performance of IB Students and Non-IB Students 24

4.2 Regression Coefficients of IB Status from Multilevel Models 29

4.3 PISA 2022 OECD country mean scores and S.D. 32
4.3.1 Mathematical Literacy 32
4.3.2 Reading 33
4.3.3 Scientific Literacy 34

References 35

List of Figures

Figure 1 Non-Modelled Performance of IB and Non-IB students in Mathematical Literacy 24
Figure 2 Non-Modelled Performance of IB and Non-IB students in Reading 25
Figure 3 Non-Modelled Performance of IB and Non-IB students in Scientific Literacy 26
Figure 4 Non-Modelled Performance of IB and Non-IB students in Narrative Writing 27
Figure 5 Non-Modelled Performance of IB and Non-IB students in Expository Writing..........................28

List of Tables

Table 1 Total IB and Non-IB Schools and Students for Current Study..4

Table 2 Performance of IB and Non-IB Students in Mathematical Literacy...9

Table 3 Performance of IB and Non-IB Students in Reading...10

Table 4 Performance of IB and Non-IB Students in Scientific Literacy...11

Table 5 Performance of IB and Non-IB Students in Narrative Writing...12

Table 6 Performance of IB and Non-IB Students in Expository Writing...13

Table 7 Comparison of Effect Sizes in Mathematical Literacy...15

Table 8 Comparison of Effect Sizes in Reading...16

Table 9 Comparison of Effect Sizes in Scientific Literacy..17

Table 10 Comparison of Effect Sizes in Narrative Writing..18

Table 11 Comparison of Effect Sizes in Expository Writing..19

Table 12 IB Schools Grade 9 and Grade 10 Student Performance in Mathematical Literacy, Reading and Scientific Literacy Relative to OECD PISA Performance 21

Table 13 Regression Coefficients of IB Status from Multilevel Models in Mathematical Literacy 29

Table 14 Regression Coefficients of IB Status from Multilevel Models in Reading...............................29

Table 15 Regression Coefficients of IB Status from Multilevel Models in Scientific Literacy 30

Table 16 Regression Coefficients of IB Status from Multilevel Models in Narrative Writing 30

Table 17 Regression Coefficients of IB Status from Multilevel Models in Expository Writing...............31

Abbreviations and Acronyms

ACER	Australian Council for Educational Research
Diff.	Difference
IB	The International Baccalaureate
ISA	International Schools' Assessment
MYP	Middle Years Programme
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PYP	Primary Years Programme
S.D.	Standard Deviation
Sig.	Significance

Executive Summary

This report presents a comparative analysis of the academic performance of students enrolled in the International Baccalaureate (IB) Primary Years Programme (PYP) and Middle Years Programme (MYP) and their non-IB peers from other accredited international schools, using data from the ACER International Schools' Assessment (ISA). The ISA is specifically designed for students in international school settings from Grades 3 to 10 and evaluates student proficiency through a combination of multiple-choice and open-ended tasks in five domains: Mathematical Literacy, Reading, Scientific Literacy, Narrative Writing, and Expository Writing. In addition, the ISA scales for Mathematical Literacy, Reading, and Scientific Literacy are aligned with the internationally recognised assessment frameworks of the OECD's Programme for International Student Assessment (PISA), allowing for robust benchmarking and international comparability.

The study uses the most recent ISA data from 2023 and 2024, to provide an updated evidence base regarding PYP and MYP student performance. Through replicating and extending the earlier methodology, this follow-up study further offers an updated evidence base and addresses two research questions requested by the IB as stated in Section 1. The analysis consists of three key components: (1) overall descriptive performance results across the five ISA domains (Section 2.1), (2) examination of the magnitude of effects in the performance differences between IB and non-IB students when accounting for the nested nature of the data with students clustered in schools (Section 2.2), and (3) benchmarking analysis of the MYP student performance on ISA against the OECD PISA 2022 performance standards (Section 2.3).

The final analytic sample includes 71,267 students from 254 schools that participated in the 2023 and 2024 ISA assessment cycles (February, May and October). All schools that were authorized to offer PYP or MYP at the time of ISA assessment cycles were included in the analysis. The comparison sample included the non-IB schools participating in the ISA, which had formal accreditation by an international education board. Of the 254 schools, there were 123 IB schools and 131 non-IB schools. Overall, the participation rates in 2023 and 2024 between IB and non-IB schools were similar across the two groups, with 44% of students coming from IB schools and 56% from non-IB schools.

To examine performance differences between IB and non-IB cohorts, ISA scale scores were analysed using two-level hierarchical linear models (HLM). Effect sizes were also calculated for all comparisons to enable direct interpretation of the magnitude of differences and comparability with non-modelled effect size estimates. Details of the methodological approach are outlined in Section 1.2 and the full results, including non-modelled estimates and outputs from different HLM models (Model 1 and Model 2), are available in Section 2.2 of this report.

A summary of Model 2 is presented in the table below. The results indicate that, in Mathematical Literacy, students significantly outperformed their peers at Grades 7 and 8, with small effect sizes. Reading was a particularly strong area for IB students, who outperformed non-IB students across almost all grades from 3 to 10, with the exception of Grade 5 where no statistically significant difference was detected. The results showed the IB students had the strongest advantage in Reading in grade 7. IB students further showed a significant advantage in Scientific Literacy at Grades 3, 4, 5, 8, and 10, with small effect sizes. Strong Reading abilities among IB students may support higher achievement in both Narrative and Expository Writing: IB students achieved significantly higher outcomes in Narrative Writing at Grades 3, 7, and 8, and in Expository Writing at Grades 3, 7, 8, and 9, with particularly strong effect (medium effect size) in grades 7 and 8.

Grade	Mathematical Literacy	Reading	Scientific Literacy	Narrative Writing	Expository Writing
3		+	+	+	+
4		+	+		
5			+		
6		+			
7	+	++		+	++
8	+	+	+	+	++
9		+			+
10		+	+		

Note: Symbols represent effect size magnitude based on multilevel Model 2 results: + small, ++ medium, and +++ large (refer to Section 1.2 for methodological details).

This project also investigated how the ISA performance of IB students in Grades 9 and 10 aligns with international benchmarks set by PISA (Programme for International Student Assessment). This comparison provides an additional perspective on the academic positioning of IB students relative to a global cohort of 15-year-olds students. The table below provides a summary of the investigation.

Grade	Mathematical Literacy	Reading	Scientific Literacy
9	+++	+++	+++
10	+++	+++	+++

Note: Symbols represent effect size magnitude based on the comparison between IB ISA mean scores and PISA 2022 OECD mean scores: + small, ++ medium, and +++ large.

The performance of Grade 9 and Grade 10 IB students were all significantly higher than the corresponding PISA 2022 OECD means with large effect sizes. In Mathematical Literacy, IB students recorded average ISA scores of 542 in Grade 9 and 563 in Grade 10. These scores are substantially higher than the PISA 2022 OECD mean of 472 for Mathematics. A similar trend is evident in Reading, where IB students achieved average scores of 532 and 552 in Grades 9 and 10, respectively, outperforming the OECD average of 476. The advantage of IB students is even more pronounced in Scientific Literacy. With average scores of 624 in Grade 9 and 633 in Grade 10, IB students were performing well above the PISA 2022 OECD mean of 485.

Although some performance advantage was observed for IB students across multiple assessment domains, these results should be interpreted with care due to several constraints in the underlying data. First, the investigation had access to only a limited set of background variables for both students and schools, restricting the ability to fully account for contextual influences. Second, participation in the ISA is voluntary rather than compulsory, and the assessment does not function as a population-level census. Therefore, the findings are representative only of institutions that chose to participate and may not reflect the broader population of international schools.

1. Project Overview

In May 2025, Australian Council for Educational Research (ACER) has accepted the request from the International Baccalaureate (IB) to investigate the performance comparison between IB and non-IB students on the International Schools' Assessment (ISA). This research report will address the following key questions as requested by IB:

- How do PYP and MYP students compare to non-IB students from schools accredited for international education on the ISA measures of mathematical literacy, reading, scientific literacy, narrative writing, and expository writing, at each grade level?
  - What is the magnitude of the effect? (IB and non-IB student performance by domain and by grade)
- How do IB students' Grade 9 and Grade 10 ISA scores compare to PISA benchmarks in each of the ISA domains that are equivalent to PISA?

The most recent ISA cycle datasets, year 2023 and 2024 sittings, are selected for answering the questions above. The ISA is an assessment designed for students in international schools from Grades 3 to 10. The program offers assessments for Mathematical Literacy, Reading, Scientific Literacy, Narrative Writing, and Expository Writing with multiple-choice and open-ended questions to provide international normative information about student performance. Especially, ISA's Mathematical Literacy, Reading, and Scientific Literacy are based on the assessment framework of OECD's Programme for International Student Assessment (PISA). Many of the international schools participating in the ISA administration implement the IB curriculum, providing an opportunity to examine the performance of IB PYP (Primary Years Programme) and MYP (Middle Years Programme) students and compare their performance to non-IB students in Grades 3 to 10.

In this report, IB students in Grades 3 to 6 are classified as PYP students and students in Grades 7 to 10 are classified as MYP students. An international school that has been participating in ISA over time may have different IB status across years. A school is a non-IB school before applying any IB authorisation. During the application, it is considered as a candidate IB school. After successfully completing the PYP and/or MYP authorisation process, it becomes an authorised IB school. Therefore, the IB status of a school at an ISA test sitting is determined by the IB authorised date, and IB programme type together with information on the ISA test sitting. For those non-IB schools, this study includes only schools formally accredited for international education.

1.1 Information on IB Schools and Students Participating in the ISA

In this study, the most recent ISA assessment data from year 2023 and 2024 were merged with the IB's data on school authorization status. After data merging, cleaning, and filtering for internationally accredited schools, the final dataset comprised 254 schools (123 IB schools and 131 non-IB schools) and 71,267 students participating in the 2023 and 2024 ISA assessment across three testing windows each year: February, May, and October. Overall, the participation rates in 2023 and 2024 between IB and non-IB schools were quite similar, with 44% of students coming from IB schools and 56% from non-IB schools. Note that the non-IB cohort consists of schools (or students in schools) with an international association but without authorised IB programme in that year/grade level. For example, students in grade 3 of an authorised MYP-only school are defined as non-IB cohort. In addition, the non-IB cohort excludes schools which are labelled as being interested in IB programmes, IB candidate schools, or schools that had withdrawn from the IB programmes. A non-IB school is defined as a school in the non-IB cohort and non-IB students are students from non-IB schools.

Table 1 outlines the number of schools and students from IB and non-IB backgrounds across Grades 3 to 10, along with the proportion of IB and non-IB students which provides an overview of the number of IB and non-IB schools and students across grades in the 2023 and 2024 sittings. The proportion of IB and non-IB schools is relatively balanced, with IB schools generally accounting for around 42 to 51% of schools in each grade. Student participation also follows a similar pattern, though non-IB students consistently make up a slightly higher share across all grades (typically from 53 to 60%), except in Grade 9 where the proportion of IB students (53%) is slightly higher. This indicates a fair distribution of participants between the two groups, but with a mild skew toward non-IB students.

Table 1 Total IB and Non-IB Schools and Students for Current Study

Grade	Current Study (2023-2024 Data)					
	Schools			Students		
	Total (N)	IB (%)	Non-IB (%)	Total (N)	IB (%)	Non-IB (%)
3	175	49	51	15249	44	56
4	160	50	50	13578	41	59
5	203	49	51	18189	45	55
6	147	46	54	11561	40	60
7	146	42	58	11899	41	59
8	130	46	54	11213	47	53
9	105	51	49	8851	53	47
10	70	49	51	4730	46	54

Note: In total, 254 schools and 71,267 students were included in the study.

1.2 Methodology

In this study, ISA scale scores were used to compare the performance of two cohorts across the five ISA domains: Mathematical Literacy, Reading, Scientific Literacy, Narrative Writing, and Expository Writing. Each domain's scale was developed using the Rasch model (Rasch, 1980) and analysed with ACER's ConQuest software (Adams et al., 2020). These scale scores were constructed by linking tests through common items (or shared questions) both within a given year (across adjacent grade levels) and across different years. This linking process allowed all tests within a domain to be equated and placed onto a single, common scale. As a result, student performance can be meaningfully compared across grade levels and over time.

To compare the performance between IB cohorts and non-IB cohorts, two-level hierarchical (multilevel) linear models (HLM) (Finch & Bolin, 2024; Hox et al., 2010; Snijders & Bosker, 2011) were employed separately for each ISA domain and grade level with the R packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) for fitting multilevel models. Restricted Maximum Likelihood (REML) estimation was employed for HLM analyses because it produces less biased and more reliable estimates of variance components than Maximum Likelihood (ML) particularly when the number of groups is small (Boedeker, 2017). After all estimates are produced, the statistical significance of the effect size of the performance differences was evaluated. In the following sections, we present our HLM models and how effect size is calculated.

1.2.1 Hierarchical (multilevel) Linear Models

HLM models are designed to account for the nested structure inherent in the ISA datasets, where individual students are clustered within schools. Failing to account for this hierarchical structure could lead to biased standard errors and incorrect inferences, due to the intra-class correlation that naturally arises from students sharing a common school environment (Zhang, 2009). This modelling approach ensures that comparisons between cohorts are robust to school-level variation and provides more reliable estimates of cohort effects across different domains and grades. Our two-level hierarchical models will be specified as follows:

1.2.1.1 Unconditional Model (Model 0)

The unconditional model is a two-level regression model which was fitted only by domain and grade.

Level 1: $Y_{ij} = \beta_{0j} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \mu_{0j}$

where:

Y_{ij} is the scale score of a domain for student *i* in school *j*,

β_{0j} is the expected average ISA score of a domain for school *j*,

e_{ij} is deviation from the expected ISA scale score of student *i* in school *j*,

γ_{00} is the grand mean of scale scores,

μ_{0j} is deviation of school *j* from the grand mean.

This unconditional model is a two-level hierarchical regression model fitted by domain and grade, without including any explanatory variables. It serves as a baseline framework for more complex conditional models as shown below.

1.2.1.2 Conditional Model 1

The conditional model 1 is a two-level regression model with IB school status in level 2.

Level 1: $Y_{ij} = \beta_{0j} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(IB_j) + \mu_{0j}$

where:

Y_{ij} is the scale score of a domain for student *i* in school *j*,

β_{0j} is the expected average ISA scores of a domain for school *j*,

e_{ij} is the deviation from the expected ISA scale scores of student *i* in school *j*,

γ_{00} is the expected average ISA scale scores for all schools,

γ_{01} is the expected difference in scale scores between IB cohort and non-IB cohort,

IB_j is a binary indicator variable for the IB status of a school *j* (1 if IB, 0 if non-IB),

μ_{0j} is the deviation of school *j* from the conditional grand mean score.

1.2.1.3 Conditional Model 2

The conditional model 2 is a two-level regression model with student gender and English-speaking background (ESB) as covariates in level 1 and IB school status in level 2.

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j}(Female_{ij}) + \beta_{2j}(ESB_{ij}) + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(IB_j) + \mu_{0j}$

where:

Y_{ij} is the scale score of a domain for student *i* in school *j*,

β_{0j} is the expected average ISA scores of a domain for school *j*,

β_{1j} is the expected difference in scale score between female students and male students,

$Female_{ij}$ is a binary indicator variable for the gender of student *i* in school *j* (1 if female, 0 if male),

β_{2j} is the expected difference in scale score between ESB students and non-ESB students,

ESB_{ij} is a binary indicator variable for the English-speaking background of student *i* in school *j* (1 if ESB, 0 if non-ESB),

e_{ij} is the deviation from the expected ISA scale scores of student *i* in school *j*,

γ_{00} is the expected average ISA scale scores for all schools,

γ_{01} is the expected difference in scale scores between IB cohort and non-IB cohort,

IB_j is a binary indicator variable for the IB status of a school *j* (1 if IB, 0 if non-IB),

μ_{0j} is the deviation of school *j* from the conditional grand mean score.

The intraclass correlation coefficient (ICC) is also used to quantifies the degree to which students within the same school resemble one another on a given outcome. In the context of two-level models (students nested within schools), the ICC is defined as the proportion of total variance that is attributable to differences between schools. The ICC or the proportion of between-school variance, is calculated as follows:

$$\text{Proportion of Between School Variance} = \frac{\text{Between School Variance}}{\text{Between School Variance} + \text{Within School Variance}}$$

A small ICC (ICC < 0.05) suggests that only a trivial proportion of total variance lies between clusters/schools, and clustering may be ignored (Muthén & Satorra, 1995). Small ICCs are usually caused by small number of schools (< 4) in either IB schools or non-IB schools. The variance estimates used to calculate the ICC are obtained from the variance components of the random effects in the unconditional model as outlined above.

1.2.2 Effect Size

Effect size offers an indication of not only whether there is a measurable difference in performance between the two groups, but also the magnitude and practical significance of that difference. In this study, the effect size Cohen's *d* (Cohen, 1988) is used to measure the magnitude of any difference. It is calculated as the difference between two means divided by a pooled standard deviation for the data:

$$d = \frac{\mu_1 - \mu_2}{\sigma_{pooled}}$$

where

$\sigma_{pooled} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$ refers to pooled standard deviation, which is the root mean square of the two standard deviations.

A key advantage of effect size is that it is independent of sample size, making it a useful metric for comparing results across groups regardless of how many participants are in each cohort. The value of the effect size provides insight into how different the IB students are from their non-IB counterparts in terms of mean ISA scale score performance. Specifically, a value close to zero suggests little to no meaningful difference in average performance between the two cohorts. A larger effect size indicates a more substantial difference in performance outcomes.

The effect size value is categorised as follows to aid interpretation.

- $d < 0.1$: Negligible difference in means. There is virtually no meaningful distinction between the two groups.
- $0.1 \le d < 0.2$: Small difference in means.
- $0.2 \le d < 0.5$: Medium difference in means. It indicates a moderate level of difference that could be of practical interest depending on the context.

- d ≥ 0.5: Large difference in means. There could be a substantial and potentially important disparity in performance.

The effect size for a regression coefficient derived from multilevel models can be calculated by standardising the coefficient relative to the variability of the outcome variable (Lorah, 2018). In this report, the regression coefficients obtained from the multilevel models are presented on the original metric of each ISA domain. We were interested in determining the expected difference in scale scores between the IB cohort and non-IB cohort which is the regression coefficient of IB (i.e. γ_{01} in the conditional model specifications above) and its effect size. To interpret this coefficient in terms of effect size, it is standardised by dividing the regression coefficient of IB by the pooled standard deviation of the outcome variable. This standardisation enables us to assess the practical significance of the estimated difference between the two cohorts, independent of the scale of measurement. The effect size also facilitates comparisons across ISA domains and supports a more general interpretation of the magnitude of the cohort effect.

The threshold for statistical significance of mean score difference and the HLM regression coefficients of IB in this study is set at a p-value of 0.05, corresponding to a 95% confidence interval. A system of symbols is used throughout the report to highlight statistically significant differences in performance between IB and non-IB schools. A single plus sign ("+") indicates that a subgroup from IB schools performed statistically significantly higher than its counterpart in the comparison group while a single minus sign ("–") denotes that a subgroup from IB schools performed statistically significantly lower than the comparison group. A double plus sign ("++") or triple plus sign ("+++") is used when the subgroup's performance is not only statistically significantly higher but also reflects a medium or large effect size. Similarly, a double minus sign ("– –") or triple minus sign ("– – –") indicates a statistically significantly lower performance with a medium or large effect size. The meanings of all symbols regarding effect size are shown below.

+	Statistically significant difference (higher), small effect size (0.1 ≤ *d* < 0.2)
++	Statistically significant difference (higher), medium effect size (0.2 ≤ *d* < 0.5)
+++	Statistically significant difference (higher), large effect size (*d* ≥ 0.5)
–	Statistically significant difference (lower), small effect size (0.1 ≤ *d* < 0.2)
– –	Statistically significant difference (lower), medium effect size (0.2 ≤ *d* < 0.5)
– – –	Statistically significant difference (lower), large effect size (*d* ≥ 0.5)

2. Analysis of Student Performance

This section addresses the two research questions mentioned in the previous section. We compared IB students (from schools authorized to offer the PYP and MYP) with non-IB students from schools accredited for international education on the ISA measures of Mathematical Literacy, Reading, Scientific Literacy, Narrative Writing, and Expository Writing, at each grade level and understand the magnitude of the effect. This section concludes with the comparison of IB students' Grade 9 and Grade 10 ISA scores to PISA benchmarks in each of the ISA domains that are equivalent to PISA.

2.1 How do PYP and MYP students compare to non-IB students from schools accredited for international education on the ISA assessment areas?

This section presents the average performance of IB and non-IB students across the five ISA domains, disaggregated by grade level. For each group, performance statistics including the number of students (N), mean scale scores (Mean), and standard deviation (S.D.) are reported alongside effect size (Cohen's *d*) of the difference between means, and the statistical significance of group differences from multilevel Model 2 (Sig. of Diff.). Slope (IB) from Model 2 is also shown in the tables to represent the adjusted mean differences after controlling for both student-level and school-level covariates. Density plots of performance of IB Students and non-IB students could be found in Appendix 4.1.

2.1.1 Mathematical Literacy

As shown in Table 2 below, IB students generally outperformed their non-IB peers in Mathematical Literacy across most grade levels. The only exception was Grade 3, where no difference was observed between the two groups.

Table 2 Performance of IB and Non-IB Students in Mathematical Literacy

Grade	IB			Non-IB			Multilevel Model 2		
	N	Mean	S.D.	N	Mean	S.D.	Slope (IB)	Effect Size	Sig. of Diff.
3	5122	350	98	7000	350	106	4.98	0.05	
4	4687	408	92	6710	407	99	3.74	0.04	
5	6667	450	89	7741	447	95	-4.79	-0.05	
6	4329	474	86	5976	472	91	-1.28	-0.01	
7	4512	508	85	6101	498	89	11.81	0.14	+
8	4764	536	95	5006	525	98	9.97	0.10	+
9	4370	542	83	3493	537	95	1.65	0.02	
10	1969	563	87	2058	560	89	1.17	0.01	

When additional student-level and school-level covariates were introduced in Multilevel Model 2, the performance gap became more pronounced in the upper grades as observed from adjusted mean differences (Slope IB), with IB students demonstrating significantly higher mean scores than non-IB students in Grades 7 and 8. These results indicate that the advantages associated with IB participation in Mathematical Literacy may become more evident as students progress through the grades. IB students continued to have slightly higher average scale scores in Grades 9 and 10; these differences, however, were negligible based on the effect size from Model 2.

2.1.2 Reading

In Reading as shown in Table 3, IB students consistently achieved higher mean scale scores than non-IB students across all grade levels. The smallest difference was observed in Grade 3 (a gap of 1 scale score), while the largest occurred in Grade 7 (a gap of 23 scale scores).

Note that while the raw mean score difference for Grade 5 was 12 (408-396) scale scores, this simply represented an unadjusted comparison that did not account for differences in student or school characteristics. Once the additional covariates were introduced in Multilevel Model 2, the adjusted difference decreased to 6.02 as shown in the slope for IB, indicating that approximately half of the original gap is explained by background factors included in the model. This illustrates how the conditional effect estimated in Model 2 reflected the difference between IB and non-IB students after controlling for relevant covariates, rather than the raw mean difference alone.

Overall, the differences in mean scale scores based on Model 2 were statistically significant from all Grades except for Grade 5, indicating a clear advantage for IB students in upper primary and lower secondary levels. The corresponding Cohen's *d* effect sizes suggest that these differences were small in Grades 3, 4, 6, and 8 to 10 but increased to a medium magnitude in Grade 7.

Table 3 Performance of IB and Non-IB Students in Reading

Grade	IB			Non-IB			Multilevel Model 2		
	N	Mean	S.D.	N	Mean	S.D.	Slope (IB)	Effect Size	Sig. of Diff.
3	5638	304	106	7981	303	110	16.65	0.15	+
4	4835	374	104	7244	367	108	13.39	0.13	+
5	7221	408	105	9187	396	107	6.02	0.06	
6	4319	440	96	6067	423	101	10.50	0.11	+
7	4571	487	98	6166	464	98	19.71	0.20	++
8	4587	515	88	5144	497	91	13.24	0.15	+
9	4344	532	86	3839	513	87	14.81	0.17	+
10	1965	552	88	2197	546	84	11.28	0.13	+

2.1.3 Scientific Literacy

In Scientific Literacy, IB students consistently outperformed their non-IB peers across all grade levels, with raw mean score differences ranging from 1 to 11 scale scores, as presented in Table 4.

The raw mean score differences at Grades 3, 4, 10 were relatively small (from 1 to 2 mean score difference), but their the adjusted differences in Model 2 increased to more than 10 points, as reflected in the slope for IB. This implied that the observed raw mean gap understated the underlying relationship between IB status and student performance. Although the magnitude of these differences varied by grade, the overall pattern indicates a steady advantage for IB students. The Cohen's *d* effect size from multilevel Model 2 analysis further revealed significant differences with small effects at Grades 3, 4, 5, 8, and 10.

Table 4 Performance of IB and Non-IB Students in Scientific Literacy

Grade	IB			Non-IB			Multilevel Model 2		
	N	Mean	S.D.	N	Mean	S.D.	Slope (IB)	Effect Size	Sig. of Diff.
3	1198	433	90	3456	431	101	14.64	0.15	+
4	1097	495	98	3245	493	98	11.41	0.12	+
5	1688	525	84	3739	514	92	11.26	0.13	+
6	1831	543	78	3865	539	86	5.11	0.06	
7	2036	563	67	2885	556	77	6.75	0.09	
8	2584	601	71	2876	591	77	9.27	0.13	+
9	1839	624	70	2389	617	81	3.47	0.05	
10	866	633	74	717	632	78	12.51	0.16	+

2.1.4 Narrative Writing

As shown in Table 5, IB students demonstrated higher mean scores than their non-IB counterparts in Narrative Writing across all grade levels, with raw differences in means lying between 1 and 12 scale scores.

The adjusted differences from IB slope in Model 2 further supported the higher performance observed in the IB group. The Cohen's *d* effect sizes showed significant differences with a small effect size in Grades 3, 7, and 8, indicating consistent advantages for IB students in these grades.

Table 5 Performance of IB and Non-IB Students in Narrative Writing

Grade	IB			Non-IB			Multilevel Model 2		
	N	Mean	S.D.	N	Mean	S.D.	Slope (IB)	Effect Size	Sig. of Diff.
3	5208	360	59	5226	359	64	6.18	0.10	+
4	4208	411	65	4507	406	70	3.85	0.06	
5	7078	453	67	6688	451	72	1.70	0.02	
6	3847	485	70	3904	482	76	1.00	0.01	
7	4137	526	72	4030	516	76	8.76	0.12	+
8	4044	559	71	3532	551	79	14.39	0.19	+
9	3921	583	78	1596	571	84	4.25	0.05	
10	1652	606	76	1437	600	80	5.98	0.08	

2.1.5 Expository Writing

In Expository Writing, IB students consistently outperformed their non-IB peers across all grade levels, as shown in Table 6, with the exception of Grade 3, where the mean scale scores were equivalent. However, the adjusted mean score difference observed from Slope IB at Grade 3 from Model 2 increased to 5.56 implying the observed raw mean gap understated the underlying relationship.

Analysis of effect sizes using Cohen's *d* from Model 2 revealed that the magnitude of differences was small in Grades 3 and 9, while a medium difference was observed in Grade 7 and 8.

Table 6 Performance of IB and Non-IB Students in Expository Writing

Grade	IB			Non-IB			Multilevel Model 2		
	N	Mean	S.D.	N	Mean	S.D.	Slope (IB)	Effect Size	Sig. of Diff.
3	5006	431	49	5213	431	57	5.56	0.10	+
4	4265	475	52	4469	471	60	2.60	0.05	
5	7048	509	55	6631	506	61	3.00	0.05	
6	3848	526	56	3853	520	64	3.45	0.06	
7	4127	556	56	3997	544	63	12.31	0.21	++
8	4066	582	58	3479	572	68	16.90	0.27	++
9	3871	597	62	1599	584	69	7.92	0.12	+
10	1578	614	60	1414	605	67	5.57	0.09	

In summary, the findings above indicate that IB students performed better than their non-IB peers in some ISA assessment domains, as evidenced by both statistical significance tests and effect size estimates (Cohen's *d*) from multilevel Model 2.

In Mathematical Literacy, IB students demonstrated higher performance than non-IB students in Grades 7 and 8 with small effect sizes. In Reading, IB students consistently outperformed non-IB students in all Grades except for Grade 5, with effect sizes ranging from small to medium, suggesting that the advantage was more pronounced in certain grade levels. For Scientific Literacy, the performance gap favoured IB students in Grades 3, 4, 5, 8 and 10 with a small effect size. In Narrative Writing, IB students achieved higher scores than non-IB students in Grades 3, 7, and 8, with small effect sizes. Finally, in Expository Writing, IB students attained significantly higher mean scores than non-IB students in Grades 3 and 7 through 9, with effect sizes varying from small to medium, highlighting a relatively stronger and more consistent performance advantage in this domain.

2.2 What is the magnitude of the effect?

This section examines the magnitude of differences between IB and non-IB students by assessing the effects of the IB school status at the school level. A series of two-level multilevel models were conducted for each domain and grade level to understand how IB school status influences student outcomes.

For each domain and grade, two conditional multilevel models were specified as detailed in section 1.2.1. Model 1 is a two-level regression model including only IB school status as a Level 2 predictor. Model 2 extends Model 1 by incorporating two student-level covariates, gender and English-speaking background (ESB), at Level 1, while retaining IB school status at Level 2. This modelling strategy allows us to account for individual-level differences while isolating the school-level. The ICC values ranging from 0.18 to 0.30 as shown in Appendix 4.2 further indicated that between school variance accounts for at least 18% of the total variance. This demonstrates a substantial clustering effect, and the magnitude of the ICC confirmed that school-level factors contributed meaningfully to the outcome differences, supporting the use of hierarchical linear models.

The magnitude of the observed effects was quantified with effect sizes, which provided a standardized metric to compare the differences between IB and non-IB students. Reporting effect sizes enables direct comparison of results derived from multilevel models with those obtained from non-modelled analyses. In the tables below, effect sizes derived from multilevel models were reported alongside the non-modelled effect size for reference and the discussion places particular emphasis on Model 2, where individual and school factors can meaningfully moderate the performance difference.

While the Model 2 is the main focus in this study, displaying the non-modelled results together with Model 1 and Model 2 provides a clear picture of how the findings change as the modelling becomes more rigorous. This full model comparison allows us to interpret the change in variance and model improvement meaningfully and trace the shifts from simple comparisons (baseline non-modelled without accounting for clustering) to the comprehensive model (Model 2). The comparison also strengthens confidence in the robustness of the findings and help explain results by comparing them from different models.

Regression coefficients of IB status from multilevel models and ICCs across all schools by domain and grade could be found in Appendix 4.2. Together, these analyses provide a comprehensive picture of how the IB students performed on the ISA in comparison with their peers across grades and domains, accounting for both individual and school level influences.

2.2.1 Mathematical Literacy

Analysis of Mathematical Literacy achievement data in Table 7 showed all three models consistently indicating that IB students performed better than their non-IB counterparts in Grades 7 and 8 with small effect sizes. For the remaining grade levels, no statistically significant differences were observed between the two groups. When additional student-level and school-level covariates were introduced in Model 1 and Model 2, the direction and magnitude of effects remained largely stable.

However, from Model 2, small negative effect sizes emerged in Grades 5 and 6 because their adjusted mean difference (IB slope) became negative once controlling for both student-level and school-level covariates. Nonetheless, these effects were not statistically significant, and the differences in mean scores were negligible, indicating that IB and non-IB students performed at comparable levels in these instances.

Table 7 Comparison of Effect Sizes in Mathematical Literacy

Grade	Non-Model		Model 1		Model 2	
	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.
3	0.00		0.05		0.05	
4	0.01		0.03		0.04	
5	0.03		-0.06		-0.05	
6	0.02		-0.03		-0.01	
7	0.11	+	0.14	+	0.14	+
8	0.11	+	0.10	+	0.10	+
9	0.06		0.01		0.02	
10	0.03		-0.01		0.01	

2.2.2 Reading

Regarding students' Reading performance (see Table 8 below), IB students demonstrated higher achievement than non-IB students in Grades 5 to 9, with small to medium effect sizes according to the non-model-based Cohen's *d*. When accounting for the nested data structure by including school-level clustering in the multilevel analysis (Model 1) and subsequently controlling for students' gender and English-speaking background (Model 2), both models consistently indicated that IB students outperformed their non-IB counterparts in Grades 3 to 10, with small to medium effect sizes, except for Grade 5.

Notably, while the non-model-based analysis revealed no statistically significant differences in Grades 3, 4, and 10, these differences became significant in the multilevel models. This implies that accounting for contextual school-level factors and individual student characteristics enhances the sensitivity of the model in detecting reading performance differences between IB and non-IB students.

Table 8 Comparison of Effect Sizes in Reading

Grade	Non-Model		Model 1		Model 2	
	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.
3	0.01		0.18	+	0.15	+
4	0.07		0.14	+	0.13	+
5	0.11	+	0.10		0.06	
6	0.17	+	0.11	+	0.11	+
7	0.23	++	0.21	++	0.2	++
8	0.20	++	0.14	+	0.15	+
9	0.22	++	0.17	+	0.17	+
10	0.07		0.14	+	0.13	+

2.2.3 Scientific Literacy

In Scientific Literacy (Table 9), IB students demonstrated higher performance than their non-IB peers, with small differences observed at Grades 5 and 8, as reflected in the corresponding Cohen's *d* values. When accounting for contextual factors at the school level in the first multilevel model (Model 1), and subsequently for individual student characteristics, including gender and English-speaking background, in the second model (Model 2), the advantage of IB students became more evident.

Under these controlled conditions, IB students continued to outperform non-IB students in Grades 3, 4, 5, 8 and 10, with small but consistent effect sizes. These results suggest that the positive association between participation in the IB programme and scientific literacy achievement persists even after controlling for both school and student level covariates.

Table 9 Comparison of Effect Sizes in Scientific Literacy

Grade	Non-Model		Model 1		Model 2	
	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.
3	0.02		0.14	+	0.15	+
4	0.02		0.11	+	0.12	+
5	0.12	+	0.15	+	0.13	+
6	0.05		0.07		0.06	
7	0.10		0.09		0.09	
8	0.14	+	0.12	+	0.13	+
9	0.09		0.05		0.05	
10	0.01		0.13	+	0.16	+

2.2.4 Narrative Writing

Regarding the Narrative Writing, results presented in Table 10, indicate that IB students consistently performed better than their non-IB peers in Grades 7 and 8, with small but meaningful differences as reflected by Cohen's *d* effect sizes across all models.

When contextual school-level factor was controlled in the first multilevel model (Model 1), and individual-level variables, including gender and English-speaking background, were further accounted for in the second model (Model 2), IB students continued to demonstrate a performance advantage over non-IB students in Grades 3, 7, and 8.

Table 10 Comparison of Effect Sizes in Narrative Writing

Grade	Non-Model		Model 1		Model 2	
	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.
3	0.02		0.13	+	0.1	+
4	0.07		0.09		0.06	
5	0.03		0.07		0.02	
6	0.04		0.02		0.01	
7	0.14	+	0.12	+	0.12	+
8	0.11	+	0.19	+	0.19	+
9	0.15	+	0.06		0.05	
10	0.08		0.09		0.08	

2.2.5 Expository Writing

The Expository Writing results displayed in Table 11 indicate that IB students consistently outperformed their non-IB peers across Grades 7 to 9, with small to medium differences as measured by Cohen's *d* effect sizes across all three models.

After accounting for school-level factors in the first multilevel model (Model 1) and further controlling for individual-level characteristics such as gender and English-speaking background in the second multilevel model (Model 2), IB students still showed a performance advantage over non-IB students, notably in Grades 3, 7, 8 and 9, with medium effect size in grade 7 and 8 and small in grades 3 and 9.

Table 11 Comparison of Effect Sizes in Expository Writing

Grade	Non-Model		Model 1		Model 2	
	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.	Effect Size	Sig. of Diff.
3	0.00		0.13	+	0.10	+
4	0.07		0.06		0.05	
5	0.05		0.08		0.05	
6	0.10		0.06		0.06	
7	0.20	++	0.20	++	0.21	++
8	0.16	+	0.26	++	0.27	++
9	0.20	+	0.13	+	0.12	+
10	0.14	+	0.09		0.09	

The results of the multilevel analysis indicate that IB students presented advantage in some domains and grade levels over non-IB peers. Strongest and most consistent gains for IB students were found in Reading, Scientific Literacy and Expository Writing across primary and secondary grades, with small to medium effect sizes based on the Model 2 results. These literacy-related skills appear to support performance in these areas. Strong Reading abilities among IB students may underpin higher achievement in both Narrative and Expository Writing.

Results for Scientific Literacy displayed grade-specific advantages that were statistically significant only after accounting for school-level and individual student factors. On the other hand, the Mathematical Literacy results revealed significant IB advantages confined to grades 7 and 8, and negligible differences in other grades.

Notably, school and individual factors play an important role in moderating these differences based on the findings from the multilevel models (Model 1 and Model 2). Some apparent advantages, particularly in early and later grades, emerge or diminish after adjusting for school-level clustering and student characteristics such as gender and English-speaking background. This implies that the performance difference between IB and non-IB students is intertwined with the broader learning environment, and that observed differences in raw scale scores from Non-Model may underestimate or overestimate the true effect of the performance difference.

It is also important to note that none of the comparisons indicated any disadvantage for IB students. Across all domains, the findings showed either no meaningful difference or a clear performance advantage for IB students relative to their non-IB peers.

2.3 How do the Grades 9 and 10 ISA scores of IB students align with PISA benchmarks?

This section presents the results of the analysis that compared IB students' Grade 9 and Grade 10 ISA scores to the OECD's PISA benchmarks in each of the ISA domains that are equivalent to PISA (for more detailed about OECD, refer to the link to the OECD website (OECD, 2025) provided in the References section). To explore how Grade 9 and Grade 10 IB students performed in comparison to international benchmarks, their ISA scores were compared against the PISA 2022 OECD mean scores. Because the ISA assessments were equated across years and grade levels and aligned with the internationally recognised assessment frameworks of the PISA, direct comparison of performance to PISA OECD mean scores within each domain was achievable. This analysis focused on Mathematical Literacy, Reading and Scientific Literacy, because no comparable PISA benchmark was available for writing tasks. It is important to note that multilevel modelling could not be applied either due to the absence of PISA microdata for hierarchical analysis.

In Table 12, it summarises the aggregated performance of Grade 9 and Grade 10 IB students in the three domains based on ISA data collected during year 2023 to 2024. Across all domains, the performance of Grade 9 and Grade 10 IB students were all significantly higher than the corresponding PISA 2022 OECD means reported by the OECD (OECD, 2023) with large effect sizes. The magnitude of the differences, expressed as Cohen's *d*, indicated large effect sizes across the board. In Mathematical Literacy, the effect sizes were 0.81 for Grade 9 and increased to 1.03 for Grade 10, indicating that IB students were performing around one pooled standard deviation above the international benchmark by the end of Grade 10. In Reading, effect sizes were slightly smaller yet still substantial, at 0.60 for Grade 9 and 0.80 for Grade 10. Scientific Literacy showed the greatest divergence from PISA OECD means, with very large effect sizes of 1.64 in Grade 9 and 1.72 in Grade 10, suggesting that IB students in these cohorts were performing more than one and a half pooled standard deviations above the OECD average.

Table 12 IB Schools Grade 9 and Grade 10 Student Performance in Mathematical Literacy, Reading and Scientific Literacy Relative to OECD PISA Performance

Domain	Grade	IB			IB vs OECD PISA Performance	
		N	Mean	S.D.	Effect Size*	Sig. of Diff.
Mathematical Literacy	9	4370	542	83	0.81	+++
	10	1969	563	87	1.03	+++
Reading	9	4344	532	86	0.60	+++
	10	1965	552	88	0.80	+++
Scientific Literacy	9	1839	624	70	1.64	+++
	10	866	633	74	1.72	+++

*Effect Size was calculated using PISA 2022 results with OECD mean =472 and S.D.=90 for Mathematics, OECD mean=476 and S.D.=101 for Reading, and OECD mean=485 and S.D.=97 for Science (OECD, 2023).

In Appendix 4.3, results also showed that Grade 9 and Grade 10 IB students continue to outperform the OECD countries' average performance across all three domains. These findings are consistent with expectations, given that PISA results reflect the performance of nationally representative samples from a wide range of school types and curricular contexts, whereas the IB cohort represents a more academically oriented and internationally focused educational pathway.

3. Conclusion

International education plays an important role in preparing students to become globally adaptable, empathetic, and competent individuals. It significantly enhances students' intercultural understanding and communication skills, while also providing strategic economic advantages for educational institutions (IB, 2025; Zhang et al., 2017). This highlights the importance of assessing whether and how different types of international education are related to observable differences in academic performance. In this context, the present study examines the performance of two groups of international students on ISA assessment to evaluate educational outcomes within diverse international learning environments.

Analysis of ISA assessment data from 2023 and 2024 revealed the consistent evidence that students enrolled in the PYP and MYP programmes outperformed their peers from international non-IB schools across multiple assessment domains. Through comparing the three sets of results (Non-Model, Model 1, and Model 2), it demonstrates the importance of Model 2, where the inclusion of both individual and school factors meaningfully moderates the observed performance differences. Based on Model 2 results, IB students demonstrated either a clear performance advantage or no meaningful difference relative to their non-IB peers. Specifically, significant differences were observed in Mathematical Literacy at Grades 7 and 8, with effect sizes exceeding 0.10. In Reading, IB students outperformed non-IB students across almost all grades from 3 to 10, except Grade 5, where no meaningful difference was detected. For Scientific Literacy, significant advantages were recorded at Grades 3, 4, 5, 8 and 10, with effect sizes greater than 0.10. The strong reading performance was also reflected on the writing domains, where IB students achieved significantly higher outcomes in Narrative Writing at Grades 3, 7, and 8, and in Expository Writing at Grades 3, 7, 8, and 9.

The positive performance trajectory of IB students was also shown in the comparative analysis using OECD PISA benchmarks across equivalent ISA domains. Across all comparable domains, the ISA mean scores of IB students were significantly higher than the PISA 2022 OECD means, with large effect sizes observed. In Mathematical Literacy, effect sizes were 0.81 at Grade 9 and increased to 1.03 at Grade 10, indicating that IB students were performing approximately one pooled standard deviation above the international benchmark by Grade 10. Reading demonstrated slightly smaller, while still noticeable, effect sizes of 0.60 at Grade 9 and 0.80 at Grade 10. Scientific Literacy showed the greatest advantage for IB students, with effect sizes of 1.64 and 1.72 for Grades 9 and 10 respectively, suggesting that IB students were performing more than one and a half pooled standard deviations above the PISA 2022 OECD average.

This study found that IB students exhibit strong performance advantages across multiple domains. However, caution needs to be used in the interpretations of these findings due to limitations in the available data. While all IB schools authorized to offer the PYP and MYP were included in this study which provides a comprehensive dataset for IB schools, ISA participation does not constitute a census, nor were the participating schools randomly sampled within each country. As participation in ISA is voluntary, findings could apply only to the sample of schools that elected to take part. In addition, at the student level, the only available background information was gender and English language background. Other potential contextual variables that the literature indicates explain variance in student achievement such as socio-economic background, school size and type, teacher qualifications and staffing ratios, funding structures, and selective enrolment policies were not available for analysis. Therefore, caution should be exercised when interpreting the results, particularly when drawing broader inferences about national or system-level performance trends.

4. Appendix

4.1 Density Plots of Non-Modelled Performance of IB Students and Non-IB Students

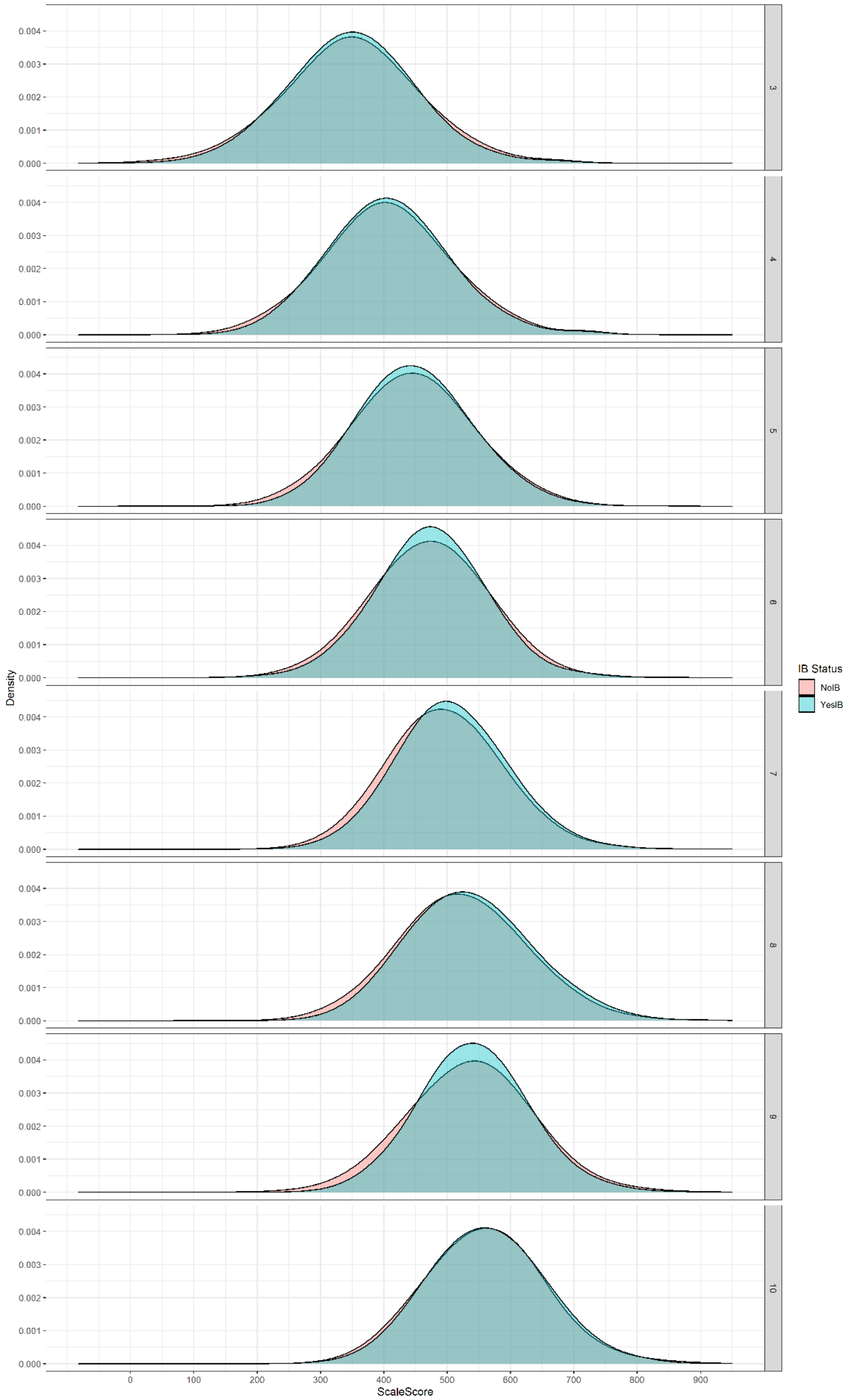

Figure 1 Non-Modelled Performance of IB and Non-IB students in Mathematical Literacy

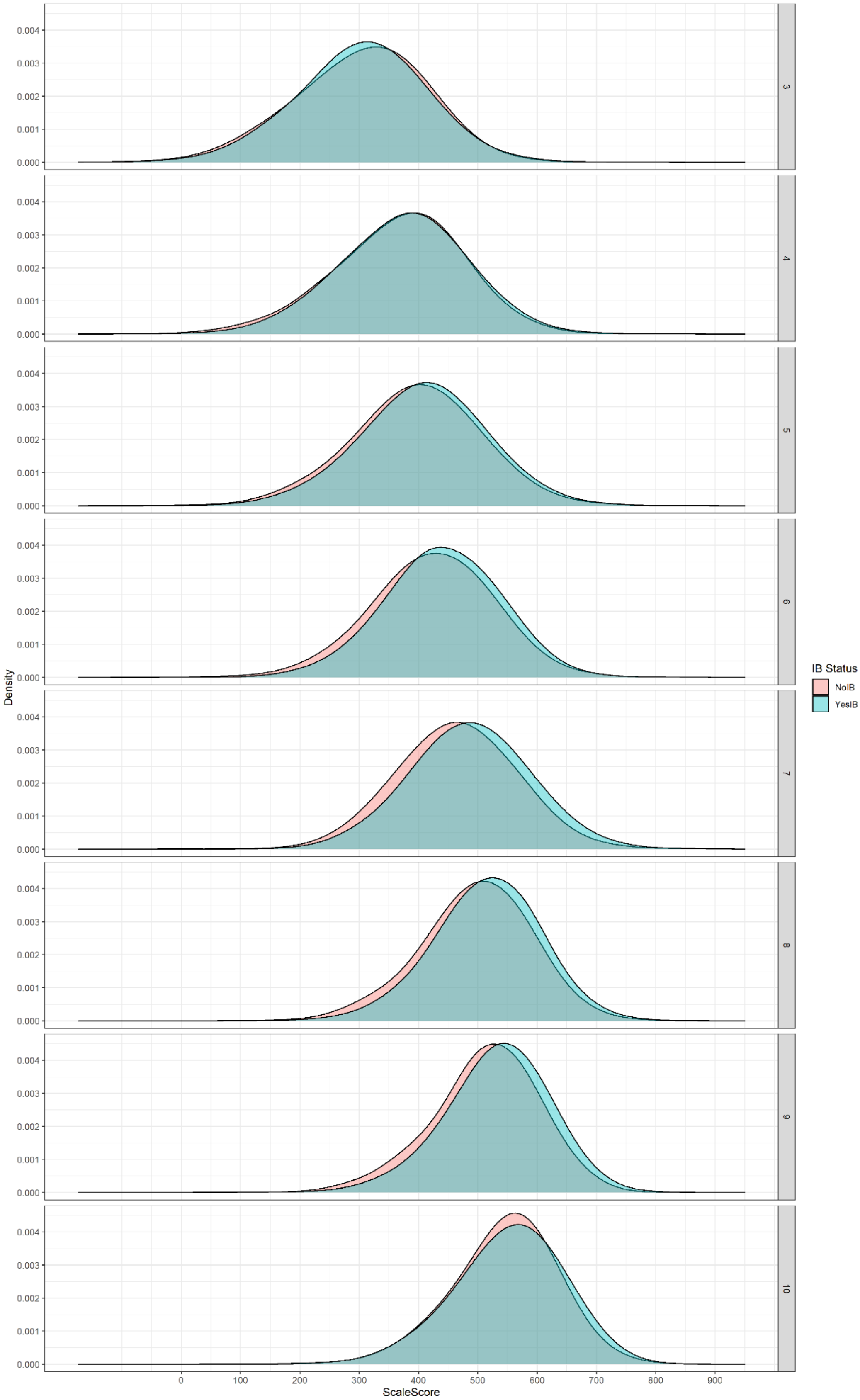

Figure 2 Non-Modelled Performance of IB and Non-IB students in Reading

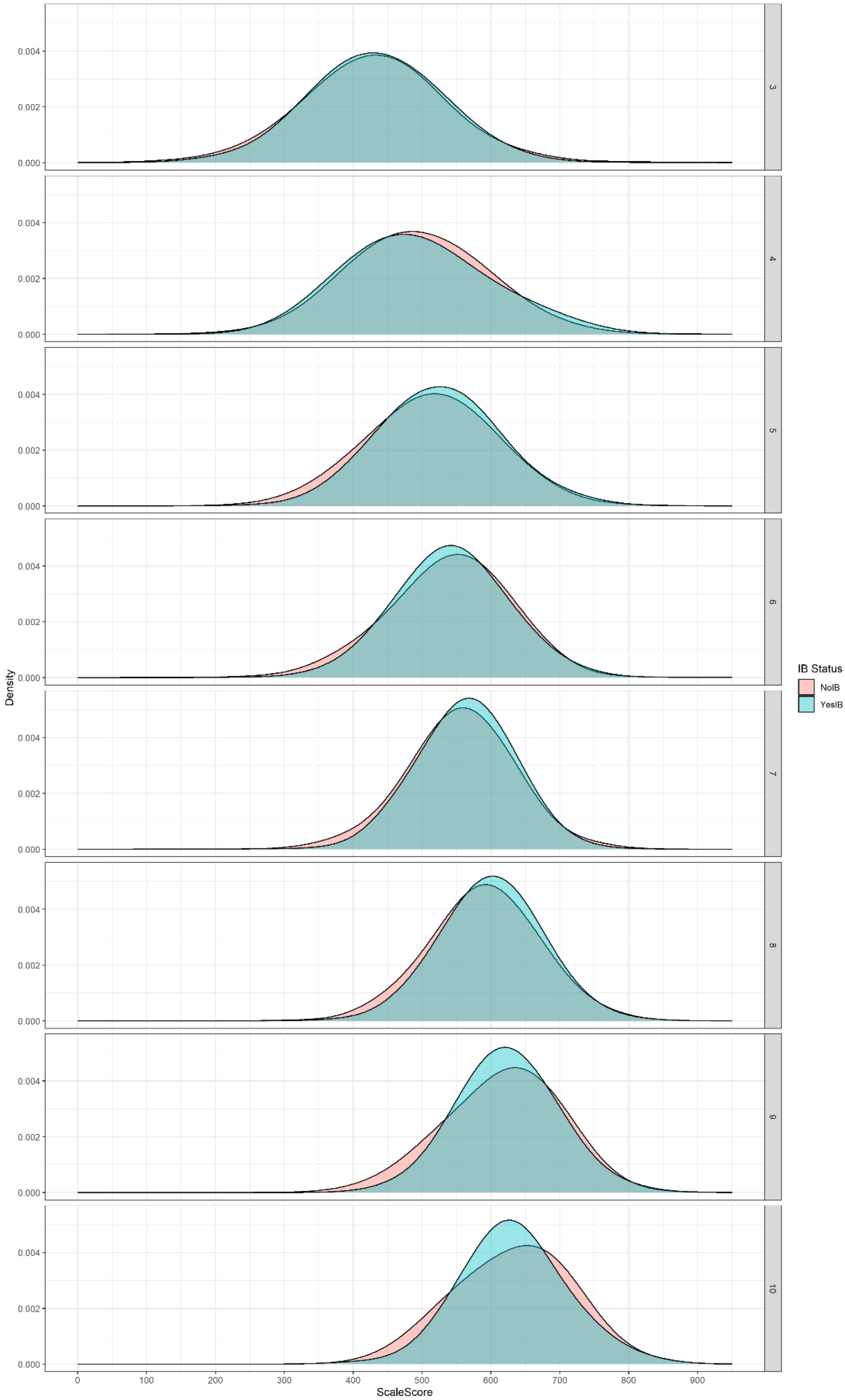

Figure 3 Non-Modelled Performance of IB and Non-IB students in Scientific Literacy

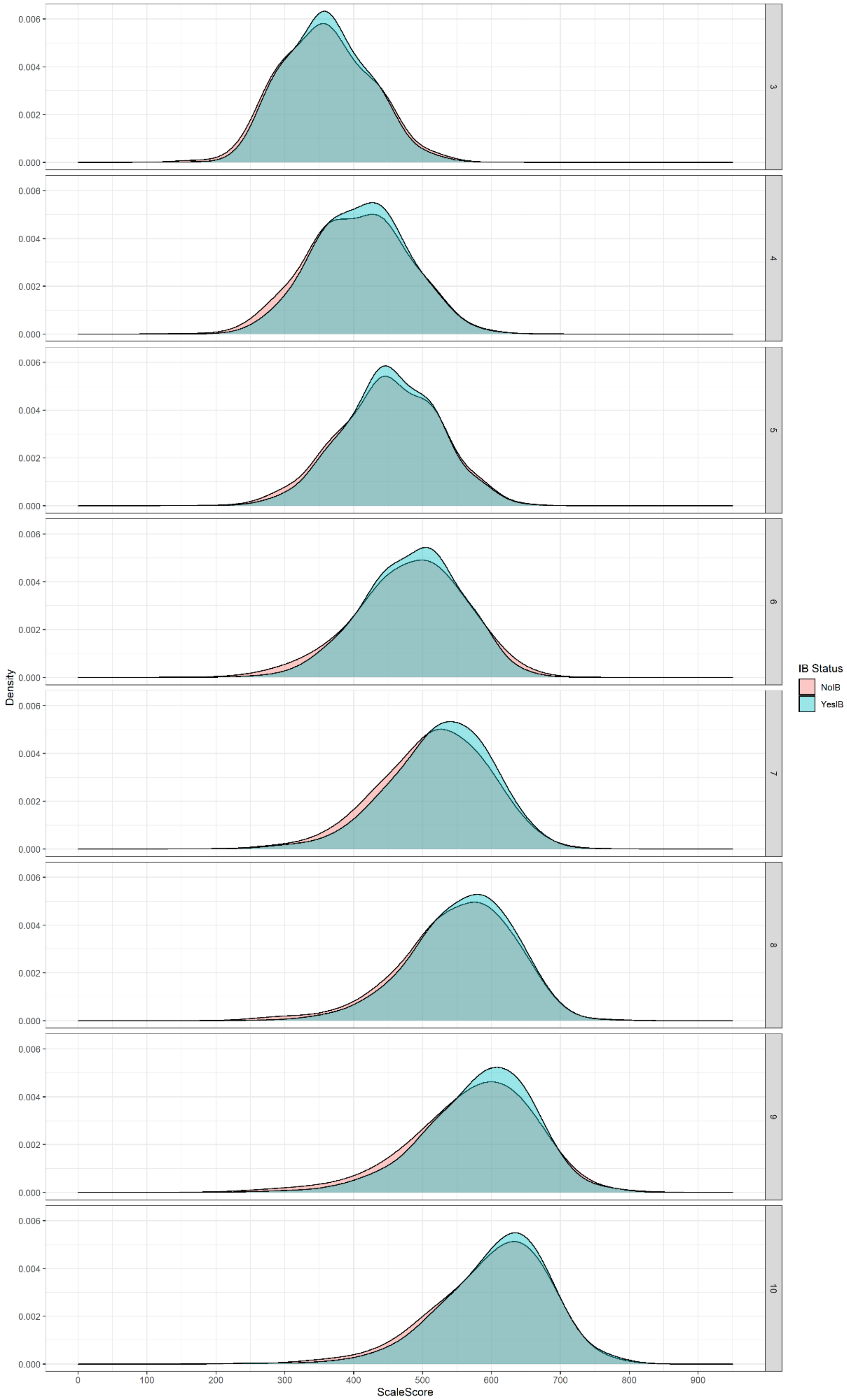

Figure 4 Non-Modelled Performance of IB and Non-IB students in Narrative Writing

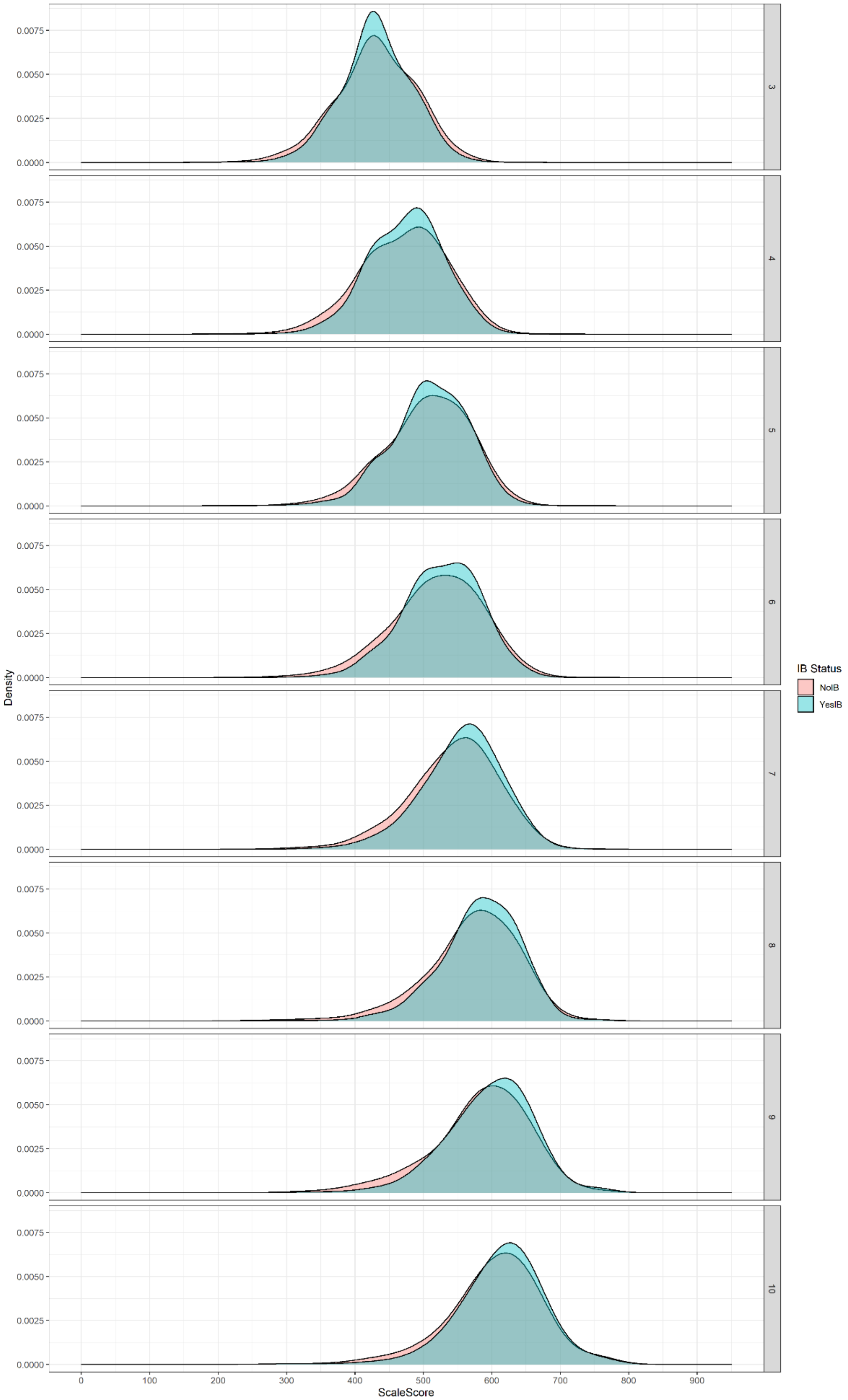

Figure 5 Non-Modelled Performance of IB and Non-IB students in Expository Writing

4.2 Regression Coefficients of IB Status from Multilevel Models

Table 13 Regression Coefficients of IB Status from Multilevel Models in Mathematical Literacy

Grade	ICC	Model 1				Model 2			
		Slope (IB)	SE	Effect Size	Sig. of Diff.	Slope (IB)	SE	Effect Size	Sig. of Diff.
3	0.22	5.29	8.05	0.05		4.98	8.01	0.05	
4	0.23	3.32	8.26	0.03		3.74	8.24	0.04	
5	0.24	-5.1	7.15	-0.06		-4.79	7.14	-0.05	
6	0.22	-2.42	7.91	-0.03		-1.28	7.83	-0.01	
7	0.20	12.05	7.24	0.14	+	11.81	7.18	0.14	+
8	0.23	10.08	9.25	0.10	+	9.97	9.23	0.10	+
9	0.20	1.02	8.65	0.01		1.65	8.58	0.02	
10	0.26	-0.75	12.10	-0.01		1.17	12.08	0.01	

Table 14 Regression Coefficients of IB Status from Multilevel Models in Reading

Grade	ICC	Model 1				Model 2			
		Slope (IB)	SE	Effect Size	Sig. of Diff.	Slope (IB)	SE	Effect Size	Sig. of Diff.
3	0.24	19.33	8.71	0.18	+	16.65	8.24	0.15	+
4	0.23	14.66	8.98	0.14	+	13.39	8.53	0.13	+
5	0.22	10.47	7.71	0.10		6.02	7.21	0.06	
6	0.23	10.95	8.84	0.11	+	10.50	8.18	0.11	+
7	0.21	20.19	8.18	0.21	++	19.71	7.52	0.20	++
8	0.22	12.95	8.41	0.14	+	13.24	7.56	0.15	+
9	0.22	14.85	8.85	0.17	+	14.81	8.25	0.17	+
10	0.22	11.63	11.03	0.14	+	11.28	10.13	0.13	+

Table 15 Regression Coefficients of IB Status from Multilevel Models in Scientific Literacy

Grade	ICC	Model 1				Model 2			
		Slope (IB)	SE	Effect Size	Sig. of Diff.	Slope (IB)	SE	Effect Size	Sig. of Diff.
3	0.23	13.76	13.69	0.14	+	14.64	13.38	0.15	+
4	0.24	11.05	14.38	0.11	+	11.41	13.86	0.12	+
5	0.20	13.36	10.77	0.15	+	11.26	10.53	0.13	+
6	0.24	5.52	10.85	0.07		5.11	10.37	0.06	
7	0.20	6.7	9.35	0.09		6.75	9.16	0.09	
8	0.21	9.09	9.13	0.12	+	9.27	8.82	0.13	+
9	0.20	4.06	11.32	0.05		3.47	10.90	0.05	
10	0.22	10.08	13.81	0.13	+	12.51	13.74	0.16	+

Table 16 Regression Coefficients of IB Status from Multilevel Models in Narrative Writing

Grade	ICC	Model 1				Model 2			
		Slope (IB)	SE	Effect Size	Sig. of Diff.	Slope (IB)	SE	Effect Size	Sig. of Diff.
3	0.21	8.13	4.97	0.13	+	6.18	4.81	0.10	+
4	0.21	5.95	5.77	0.09		3.85	5.50	0.06	
5	0.18	4.81	4.80	0.07		1.70	4.47	0.02	
6	0.21	1.18	6.64	0.02		1.00	5.93	0.01	
7	0.22	8.57	6.82	0.12	+	8.76	6.08	0.12	+
8	0.23	14.36	7.47	0.19	+	14.39	6.67	0.19	+
9	0.22	5.21	8.88	0.06		4.25	8.04	0.05	
10	0.24	7.04	11.18	0.09		5.98	9.84	0.08	

Table 17 Regression Coefficients of IB Status from Multilevel Models in Expository Writing

Grade	ICC	Model 1				Model 2			
		Slope (IB)	SE	Effect Size	Sig. of Diff.	Slope (IB)	SE	Effect Size	Sig. of Diff.
3	0.25	6.83	4.71	0.13	+	5.56	4.62	0.10	+
4	0.27	3.64	5.48	0.06		2.60	5.38	0.05	
5	0.24	4.70	4.57	0.08		3.00	4.39	0.05	
6	0.25	3.60	5.98	0.06		3.45	5.62	0.06	
7	0.25	11.98	5.83	0.20	++	12.31	5.53	0.21	++
8	0.25	16.67	6.47	0.26	++	16.90	6.12	0.27	++
9	0.24	8.22	7.35	0.13	+	7.92	6.91	0.12	+
10	0.30	5.98	10.24	0.09		5.57	9.61	0.09	

4.3 PISA 2022 OECD country mean scores and S.D.

4.3.1 Mathematical Literacy

IB performance / OECD Country	Mean Score	S.D.
IB Grade 9 on the ISA	542	83
IB Grade 10 on the ISA	563	87
OECD average	472	90
Australia*	487	99
Austria	487	94
Belgium	489	96
Canada*	497	94
Chile	412	77
Colombia	383	73
Costa Rica	385	66
Czech Republic	487	93
Denmark*	489	82
Estonia	510	85
Finland	484	89
France	474	91
Germany	475	95
Greece	430	83
Hungary	473	94
Iceland	459	88
Ireland*	492	80
Israel	458	107
Italy	471	89
Japan	536	93
Korea	527	105
Latvia*	483	80
Lithuania	475	87
Mexico	395	69
Netherlands*	493	106
New Zealand*	479	99
Norway	468	93
Poland	489	89
Portugal	472	90
Slovak Republic	464	101
Slovenia	485	89
Spain	473	86
Sweden	482	96
Switzerland	508	96
Türkiye	453	90
United Kingdom*	489	96
United States*	465	95

* Caution is required when interpreting estimates because one or more PISA sampling standards were not met (see Reader's Guide, Annexes A2 and A4) (OECD, 2023).

4.3.2 Reading

IB performance / OECD Country	Mean Score	S.D.
IB Grade 9 on the ISA	532	86
IB Grade 10 on the ISA	552	88
OECD average	476	101
Australia*	498	111
Austria	480	104
Belgium	479	105
Canada*	507	109
Chile	448	93
Colombia	409	93
Costa Rica	415	86
Czech Republic	489	98
Denmark*	489	92
Estonia	511	92
Finland	490	104
France	474	106
Germany	480	106
Greece	438	94
Hungary	473	101
Iceland	436	103
Ireland*	516	88
Israel	474	122
Italy	482	92
Japan	516	96
Korea	515	103
Latvia*	475	90
Lithuania	472	94
Mexico	415	84
Netherlands*	459	115
New Zealand*	501	109
Norway	477	112
Poland	489	104
Portugal	477	94
Slovak Republic	447	105
Slovenia	469	97
Spain	474	97
Sweden	487	111
Switzerland	483	105
Türkiye	456	87
United Kingdom*	494	105
United States*	504	111

* Caution is required when interpreting estimates because one or more PISA sampling standards were not met (see Reader's Guide, Annexes A2 and A4) (OECD, 2023).

4.3.3 Scientific Literacy

IB performance / OECD Country	Mean Score	S.D.
IB Grade 9 on the ISA	624	70
IB Grade 10 on the ISA	633	74
OECD average	485	97
Australia*	507	109
Austria	491	101
Belgium	491	101
Canada*	515	101
Chile	444	92
Colombia	411	87
Costa Rica	411	80
Czech Republic	498	99
Denmark*	494	95
Estonia	526	89
Finland	511	106
France	487	103
Germany	492	106
Greece	441	91
Hungary	486	96
Iceland	447	95
Ireland*	504	91
Israel	465	109
Italy	477	93
Japan	547	93
Korea	528	105
Latvia*	494	85
Lithuania	484	92
Mexico	410	75
Netherlands*	488	112
New Zealand*	504	107
Norway	478	106
Poland	499	96
Portugal	484	92
Slovak Republic	462	103
Slovenia	500	94
Spain	485	92
Sweden	494	108
Switzerland	503	99
Türkiye	476	89
United Kingdom*	500	104
United States*	499	108

* Caution is required when interpreting estimates because one or more PISA sampling standards were not met (see Reader's Guide, Annexes A2 and A4) (OECD, 2023).

References

Adams, R. J., Wu, M. L., Cloney, D., Berezner, A., & Wilson, M. R. (2020). *ACER ConQuest 5*. Australian Council for Educational Research. https://www.acer.org/au/conquest

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Boedeker, P. (2017). Hierarchical linear modeling with maximum likelihood, restricted maximum likelihood, and fully Bayesian estimation. *Practical Assessment, Research, and Evaluation*, *22*(1), 1-19. https://doi.org/10.7275/5vvy-8613

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Routledge.

Finch, W. H., & Bolin, J. E. (2024). *Multilevel modeling using R* (3rd Ed.). CRC Press.

Hox, J., Moerbeek, M., & Van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications* (2nd Ed.). Routledge.

IB. (2025, March 06). *International education: it's time to think again.* IB World archive. https://ibo.org/ib-world-archive/september-2012/international-education-its-time-to-think-again

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-scale assessments in education*, *6*(8), 1-11. https://doi.org/10.1186/s40536-018-0061-2

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267-316. https://doi.org/10.2307/271070

OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Organisation for Economic Co-operation and Development. https://doi.org/10.1787/53f23881-en

OECD. (2025). *The OECD: Better policies for better lives*. https://www.oecd.org/en/about.html

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.

Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd Ed.). Sage Publishing.

Zhang, L.-C. (2009). *The Aftereffect of Higher Education Expansion Policy - Over-education, evidence from HLM analysis.* In H. L. Pan, (Chair), Dialogue between Educational Research and Educational Policy, Symposium conducted at the meeting of international conference, National Taiwan Normal University, Taiwan.

Zhang, L.-C., Worthington, A. C., & Hu, M. (2017). Cost economies in the provision of higher education for international students: Australian evidence. *Higher Education*, *74*(4), 717-734. https://doi.org/10.1007/s10734-016-0078-9