# Predictors of Academic Outcomes in the International Baccalaureate Diploma Programme

Final Report

Kit S. Double, Therese N. Hopfenbeck & Joshua A. McGrane

Oxford University Centre for Educational Assessment

DEPARTMENT OF EDUCATION

UNIVERSITY OF OXFORD

OUCEA

Oxford University Centre for Educational Assessment

## Acknowledgements

We would like to thank Dr Olivia Halic for her support and input throughout the research project, and to the International Baccalaureate and researchers at the Department of Sociology of VU University Amsterdam who were responsible for the original data collection, as well as the IB Diploma Programme schools, teachers and students who provided the original data.

# Table of Contents


Executive Summary ..... 4
Context ..... 4
Scope and objectives ..... 4
Methodological approach ..... 4
Main findings ..... 4
Recommendations ..... 5
Concluding statement ..... 5
Introduction ..... 6
Background ..... 6
Research Aims and Rationale ..... 7
Methodology ..... 8
Participants ..... 8
Missing data ..... 8
Survey Instruments ..... 8
Academic Performance ..... 12
Analysis and Findings ..... 13
Survey Scale Reliability ..... 13
Academic Performance ..... 14
Variance Decomposition ..... 19
Student Level Models ..... 19
School Level Models ..... 28
Interactive Models ..... 31
Change Models ..... 38
Discussion ..... 40
Recommendations ..... 42
Conclusions ..... 42
References ..... 43
Appendices ..... 46
Appendix A ..... 46
Appendix B ..... 48
Appendix C ..... 50

# Executive Summary

## Context

This report outlines the analyses that have been completed by the Oxford University Centre for Educational Assessment (OUCEA) concerning the predictors of academic success during the IB Diploma Programme (DP). The data analysis is a secondary analysis of the longitudinal survey of DP students' experience and outcomes, which was administered by IB Research during 2016–2018. The primary aim of the data analysis is to use student and DP coordinator survey responses to predict a measure of academic performance that is derived from scaled exam grades from the May 2018 session and grades for the 2018 Theory of Knowledge and Extended Essay subjects.

## Scope and objectives

The data is analysed with four main goals: (1) identify the student-level and school-level factors that predict DP academic performance, (2) estimate the relative importance of each of the identified factors, (3) assess the cross-level interactions between school-level variables and student-level variables to understand how these different levels combine to affect student outcomes, and (4) estimate the variance in academic performance that occurs at the between school-level compared to the within-school level.

## Methodological approach

Student-level survey data was collected in three waves in Diploma Programme schools and covered a broad range of social, contextual, and psychological variables. This data is complemented with survey data collected from DP co-ordinators as part of the original data collection procedure. In total the data concerns 4,858 students from 99 schools, in 36 countries.

The main analyses included in the report are multilevel regression models using Stepwise Regression (Chatterjee & Hadi, 2012). This stepwise procedure began with complete models using all variables which were then reduced using model selection based on the Akaike Information Criterion (AIC). The final models represent relatively parsimonious models of the best predictors of the academic performance.

## Main findings

The following summarises the key findings of the analyses:

• Students **self-reported grades** are consistently the best predictors of academic performance.

• **Exam preparation, studying, and completing homework** consistently positively predict academic performance.

• Activities that place additional **time-demands** on students, such as doing chores and working for pay, negatively predict academic performance.

• Students who became either happier or less isolated across the course of the DP tended to perform better academically.

• While the bivariate relationships between academic performance and individual predictors are relatively small, overall, the models explain a substantial proportion of variance in academic performance (typically around 50%).

• Significant **school-level predictors** related to homework and assessment practices, but also pointed to the importance of school composition and size (e.g., higher number of DP staff members and coordinated IA dates both predict better academic performance).

• Approximately **42% of the variance** in academic performance occurred between schools, while **58% occurred** within schools. This suggests that both school-level and student-level variables are likely to make important contributions in determining students' academic performance.

## Recommendations

The findings here suggest that academic performance is driven by the cumulative effect of many student and school characteristics. It is therefore prudent that the IB continue to focus on developing a wide range of student characteristics, especially helping students manage their time and activities, both academic and non-academic, inside and outside of the classroom. While these findings are largely unsurprising, it is comforting to consider that the best preparation for DP examination appears to be the time that students spent inside and outside of class learning, preparing, and revising. The research also indicates that promoting increases in students' happiness and reducing their sense of isolation over the course of the DP may be a promising means of improving their academic performance.

## Concluding statement

Overall, the findings presented in this report indicate the role of a diverse set of student and school factors in predicting academic performance in the DP. The findings align well with previous research and suggest that a comprehensive understanding of a multitude of student and school variables, along with their interactions, is necessary when designing educational interventions and policy reforms to improve academic performance. The findings suggest that the strongest predictor of academic performance is students' self-reported academic achievements along with factors related to instructional and learning time, such as class time, study time, and not having to do household chores and work for pay. Moreover, changes in happiness and feelings of isolation throughout the duration of the DP were shown to be associated with final performance. Future research should endeavour to measure more variables over time to allow for more extensive longitudinal analyses, including more indicators of students' home and school context, to develop a temporally richer and well controlled model of predictors of DP academic outcomes.

# Introduction

## Background

Research has highlighted the importance of understanding the long-term predictors of academic outcomes and the role of both student-level and school-level factors in predicting these outcomes (Hattie, 2014). Importantly, understanding how school policies/context interact with student characteristics is important for developing effective school policies and promoting equity (Leckie, 2009).

The nature and sources of individual differences in learning and academic performance is the subject of considerable scientific discussion and large-scale research syntheses have been carried out to examine the reliability and relative importance of student and school factors for academic performance (e.g., Hattie, 2014). These syntheses have suggested that there are many factors that predict students' academic success, some of which are related to students' background characteristics (socioeconomic status, motivation, intelligence, etc. Sirin, 2005), and others to school characteristics (e.g., staff to student ratios; Hattie, 2014). Moreover, research has suggested that up to 80% of the variance in academic achievement can be accounted for by student background and school conditions (Chen & Weikart, 2008). Many of the recent findings have also stressed the importance of non-cognitive factors in predicting academic success (e.g., Lee & Stankov, 2018; Stankov, 2013). Some of the most important student-level variables that have been correlated with academic success include engagement, self-concept, anxiety, self-regulation (Hattie, 2014; Richardson, Abraham, & Bond, 2012) and parental involvement (Fan, 2001). Moreover, a recent large-scale meta-analysis suggested that emotional regulation (EQ) was almost as important for predicting academic success as cognitive abilities (MacCann et al., 2020).

It is also important to consider the influence of school policy and context on academic performance, as a wide range of school-levels variables have been shown to impact of academic performance, including school climate (Haynes, Emmons, & Ben-Avie, 1997; Uline & Tschannen-Moran, 2008), school wealth and type (independent vs. public funded; Thiele, Singleton, Pope, & Stanistreet (2016)), and classroom size (Werblow & Duesbery, 2009). Additionally, previous research has rarely simultaneously considered the effects of both student and school variables on academic outcomes. This is particularly important because educational outcomes are often the product of complex interactive effects resulting from a combination of both school and student effects (Kwok et al., 2018).

Recent research with IB students has examined the predictors of academic achievement in the DP, including its relationship with prior achievement, percentage of second language learners in a school, and gender (Ballantyne & Rivera, 2014). While these studies provide an initial insight into the relationship between school context, student characteristics, and academic outcomes in the IB, to date, there has been no in-depth modelling study of the interactive effects of a broad range of school and student characteristics on DP outcomes. This project expands on these recent findings using multilevel regression modelling techniques to examine the predictors of academic outcomes within the DP programme.

While research with IB students has largely focused on whether DP performance predicts outcomes in higher education (e.g., Halic, 2013; Pilchen, Caspary, & Woodworth, 2019), recent research has begun to explore the factors that do and do not predict DP performance. For instance, Ballantyne and Rivera (2014) found that factors like gender and the percentage of second-language learners within a school did not predict academic outcomes in the DP.

## Research Aims and Rationale

Overall, the report aims to describe the best predictors of academic performance in the DP at both the student and school-levels, as well as to identify interactions between student and school variables that predict academic outcomes. To accomplish these goals, we used multilevel models to analyse the data. The approach allowed us to examine how each of the relevant predictor variables relates to academic performance at both the school and student level. Furthermore, we examine the cross-level interactions to assess whether school policies and context are more impactful for some students compared with their peers. In addition, we examine whether changes in student variables across the course of the DP predict academic outcomes. Finally, an advantage of this methodology is that we can estimate the variance that occurs at the student and school levels, which can provide important evidence about the relative contribution of student and school-level effects on academic performance.

The data is analysed with four main goals: (1) identify the student-level and school-level factors that predict DP academic performance, (2) estimate the relative importance of each of the identified factors, and (3) assess the cross-level interactions between school-level variables and student-level variables to understand how these different levels combine to affect student outcomes, and (4) estimate the variance in academic performance that occurs at the between school-level compared to the within-school level.

# Methodology

## Participants

The data were originally collected by the IB as part of their longitudinal survey of DP students' experience and outcomes during 2016–2018. The focus of this original survey was to evaluate how the workload demands in the Diploma Programme affect student wellbeing. Student data was collected in three waves (beginning of year 1 of the DP, end of year 1 of DP and at the completion of the programme, after the May 2018 exam session) from 4,858 students enrolled in 99 schools in 36 countries. Student survey responses were matched to the May 2018 IB exam results, including IB subject grades and Extended Essay and Theory of Knowledge course performance.

Additionally, complimentary school-level data was collected through an online questionnaire distributed to DP co-ordinators at the beginning of year 2 of the DP. In total, 91 schools provided survey responses from a DP co-ordinator. Twenty-eight schools submitted survey data from multiple DP co-ordinators, so the second set of responses were removed from the data. The sampling procedure ensured that the school sample is representative of the IB school population regarding geographical distribution, language of instruction, number of DP schools by country and school status (private/state-funded).

## Missing data

To handle missing data, we utilised a random forest imputation algorithm, which assumes that data was missing at random (Stekhoven & Bühlmann, 2012). Random forest imputation is a machine learning approach to missing data that uses a non-parametric imputation method applicable to various variable types (i.e., it does not make any distributional assumptions). This approach has been shown to work well for ordinal missing data (Stekhoven & Bühlmann, 2012). Several questions were flagged by the researchers as unlikely to be missing at random because, for example, missing data was likely to represent that the question did not apply to the respondent. For such cases, an imputed score indicating not applicable was used. Missing data in the DP co-ordinator survey was handled in the same fashion as the student survey data.

## Survey Instruments

The survey data was collected in three waves. The waves differed substantially from each other in terms of the measures/scales included, but several items/scales were repeated across the three waves. The survey instruments are available upon request by emailing research@ibo.org. The surveys included several single-item measures of pertinent variables as well as a number of multi-item scales. Data concerning several demographic variables (e.g., subject choice) was collected, but not included in the analysis. For brevity, we discuss the multi-item scales here, and any interested reader should contact the IB (research@ibo.org) for the complete survey instruments.

## Student-level Scales

### DP Connectedness

The DP connectedness scale was a 4-item scale that assessed how connected students felt to their peers and the DP programme (e.g., "I feel that DP students care about each other."). Responses were on a 4-response Likert-like scale ranging from *All of the time* to *None of the time*. This scale was administered in Wave 1 only.

### DP Demands

The DP demand scale was a 3-item scale that evaluated how students viewed the pressure and demands of the DP programme (e.g., "There is much pressure in the DP to excel."). Responses were on a 4-response Likert-like scale ranging from *Strongly disagree* to *strongly agree*. This scale was administered in Wave 1 only.

### DP Feedback

The DP feedback scale asked 4-items concerning the extent to which students felt they received feedback or felt they could ask questions in the DP (e.g., "It is hard to get help in my DP classes when I have a question"). Responses again ranged from *Strongly disagree* to *Strongly agree*. This scale was administered in Wave 1 only.

### DP Sentiment

The DP sentiment scale asked students how they felt about three aspects of the DP (subjects, teachers, and classmates). Each item began with "How do you feel about…?" and was rated a 6-response Likert-like scale ranging from Completely *dissatisfied* to *Completely satisfied*. This scale was administered in Wave 1 only.

### Happiness

Student's happiness was assessed using a 3-item scale that asked about students' happiness in general, a year ago, and how happy they expected to be in a year's time on a 7-point scale ranging from *Extremely unhappy* to *Extremely happy*. This scale was administered in Waves 1 and 2 only.

### Isolation

The 3-item isolation scale asked students to report how socially isolated they feel (e.g., "I feel that there is no one I can share my most private worries and fears with.") on a 4-point scale from *Definitely false* to *Definitely true*. This scale was administered in Waves 1-3.

### Lack of Control

The 10-item lack of control scale assessed students' control over difficulties in their life. The scale was prefaced with "During the past four weeks, how often have you…" and participants responded to each item on a 5-point scale ranging from *All of the time* to *None of the time and*

*Not Applicable* (e.g., "…felt confident about your ability to handle your personal problems?"). This scale was administered in Waves 1-3.

### Language Understanding

The language understanding scale asked students to rate their overall language ability on five aspects on language (e.g., reading). Each item was rated on a 5-point scale ranging from *Poor* to *Native speaker command*. This scale was administered in Waves 1 and 2 only.

### Life Satisfaction

The 5-item asked students to rate how their life is going (e.g., "My life is going well.") on a 6-point scale ranging from *Strongly disagree* to *Strongly agree*. This scale was administered in Waves 1-3.

### Absenteeism

6-items asked students to report how often they missed school in the last two weeks. Three items were framed in terms of absenteeism related to illness, while three items asked students about absenteeism due to not wanting to attend school. Response options ranged from *Never* to *5 times or more*. This scale was administered in Waves 1 and 2 only.

### Organisation

The 5-item organisation scale asked students about their organisation and time management (e.g., "When you are (or have been) faced with school-related challenges, how often do you purposely think about how to schedule and spend your time?"). Students rated each item from *All of the time* to *None of the time* on a 5-point scale. This scale was administered in Waves 1 and 2 only.

### Parental Involvement

The 15-item parental involvement scale measured students' parental involvement by asking about their parents' involvement in various aspects of their school-life (e.g., "During the last week when you were in school, about how often did one of your parents check your homework after it was completed?"). Five items involved students making overall judgments about their parents' involvement on a 5-point scale from *Not at all involved* to *Extremely involved*. Six items asked students about their parents' involvement in the last week/month and were rated on 4-point scales from *Never* to *More than four times*. Finally, four items asked about parental expectations and were rated on a 4-point scale from *Strongly disagree* to *Strongly agree*. This scale was administered in Waves 1 and 2 only.

### Time Poorness

The 5-item time poorness scale asked participants about their energy and accomplishments over the past four weeks (e.g., "During the past four weeks, how much of the time did you accomplish less than you would have liked?") on a 5-point scale from *All of the time* to *None of the time*. This scale was administered in Waves 1-3.

### Worry

A 5-item worry scale assessed the extent to which students worry about aspects of their school performance (e.g., "I worry that I will get poor grades at school."). Responses were on a 5-point scale ranging from *All of the time* to *None of the time*. This scale was administered in Waves 1 and 2 only.

### Stress

A 3-item stress scale asked students how stressed they were on average at different time-points during their DP year (e.g., "On average, how stressed would you say you were during your second year in the DP? - December - February"). Responses were on a 5-item scale from *Not at all stressed* to *Extremely stressed*. This scale was administered in Wave 3 only.

## School-level Scales

### Wellbeing Staff

A three-item scale asked staff to rate whether different wellbeing related staff were available at the school. Items referred to a student wellbeing team, a staff wellbeing team, and an internal welfare coordinator. Items were rated Yes/No based on their availability in the school.

### Disruption

A 7-item scale assessed the extent to which students' learning is disrupted by various phenomenon (e.g., absenteeism, poor student-teacher relations). Each item was rated on a 4-point scale from *Not at all* to *To a great extent*.

### Support

The support scale was assessed using a check-list response where staff rated which of a range of possible support services were available in their school (e.g., teacher-led subject specific tutoring, university counselling, summer school). A total score was computed by summing the number of available support services.

### Parental Involvement

Staff rated the involvement and expectations of their schools' parents on a 14-item scale. Items assessed the extent to which the school welcomes parental input, the percentage of parents who get involved in school activities, parents' expectations, and parental pressure.

### DP Connectedness

Staff rated the extent to which DP students were connected to each other and the sense of caring and spirit amongst the DP cohort (e.g., *I feel that DP students care about each other*) on a 4-item scale with response options ranging from *All of the time* to *None of the time*.

## Academic Performance

Students' overall academic performance was calculated using an Item Response Theory (IRT) approach. In this IRT approach, each of the students' DP subject grades along with their Extended Essay and Theory of Knowledge grades were included as 'items'[1] and their ability estimate from the model was used as the outcome variable of academic performance in the predictive modelling. Using an IRT approach provides academic performance estimates that take account of differences in difficulty across the different DP components, as well as potential differences in the degree to which the components discriminate between high and low academic performers.

Both the Partial Credit (PCM) and Generalized Partial Credit Models (GPCM) were applied to and compared for the grade data in terms of fit[2]. The former is a simpler model and therefore provides more robust estimates for the DP subjects with small sample sizes in the data, and the latter allows for differences in discrimination for the different components, which is important, as subjects with a larger proportion of coursework (or total in the case of the Extended Essay) are known to discriminate less between low and high performers, i.e., only a small range of changes in grades is observed over a large range of abilities as estimated through performance across all subjects. The model fit comparison was done based on the Akaike Information Criterion (AIC; Chatterjee & Hadi, 2012). The AIC is typically used in model comparison and is a measure of prediction error (with smaller values indicating better model fit). All IRT analyses were conducted using the TAM package in R (Robitzsch, Kiefer, & Wu, 2020).

---

[1] Each DP subject code (see Table 11 in Appendix C) was treated as a unique 'item' and so the response matrix was sparsely populated, however, this kind of missing data is routinely handled by IRT modelling and the difficulty and ability estimates are linked by the overlap between different students taking the same subjects in differing combinations.

[2] Both the PCM and GPCM are IRT models that are applicable to polytomous data, i.e., data where the items (in this case, the subjects) have more than two possible score categories, which is the case for the IB DP subject grades. They both model the item scores as a trade-off relationship between the difficulty of the item, the difficulty of each score category, and the ability (theta estimate) of the student. So, the higher a student's ability, as estimated in terms of their total score across all subjects, relative to the difficulty of an item, the more likely they are to score in a higher category for that item, and vice-versa. In addition, the GPCM allows each item to discriminate differently, whereas the PCM constrains the discrimination to be equal across items. Discrimination relates to the required range of ability to go from the lowest category being the most probable score to the highest category being the most probable for an item. Lowly discriminating items correspond to a large range of abilities being required to go from the bottom to the top category being the most probable score, and vice-versa for highly discriminating items. Coursework grades are often found to be less discriminating than exam grades, as in the case of the former, only extremely lowly performing students achieve the lowest score categories and often only the most extremely highly performing students achieve the highest score category. This pattern can be observed in the category threshold difficulty (tau) estimates for the Extended Essay and Theory of Knowledge subjects in Table 10 in Appendix B.

# Analysis and Findings

## Survey Scale Reliability

The reliability of the included scales, assessed using Cronbach's alpha, is presented in Table 1. Reliability varied substantially across the scales, with most scales only having moderate reliability. Intercorrelations between the scales are shown in Appendix A.

*Table 1. Reliability (Cronbach's alpha) of student-level scales*

| Scale | Wave 1 | Wave 2 | Wave 3 |
|---|---|---|---|
| **Student-level Scales** | | | |
| *DP connectedness* | 0.826 | NA | NA |
| *DP Demands* | 0.760 | NA | NA |
| *DP feedback* | 0.701 | NA | NA |
| *DP sentiment* | 0.671 | NA | NA |
| *Happiness* | 0.722 | 0.659 | 0.561 |
| *Isolation* | 0.677 | 0.604 | 0.620 |
| *Lack of Control* | 0.889 | 0.550 | 0.466 |
| *Language understanding* | 0.959 | 0.969 | NA |
| *Life satisfaction* | 0.882 | 0.871 | 0.854 |
| *Absenteeism* | 0.743 | 0.810 | NA |
| *Organising* | 0.818 | 0.761 | NA |
| *Parental involvement* | 0.803 | 0.348 | NA |
| *Stress* | NA | NA | 0.387 |
| *Time poorness* | 0.798 | 0.786 | 0.772 |
| *Worry* | 0.945 | 0.883 | NA |
| **School-level Scales** | | | |
| *Wellbeing Staff* | 0.757 | | |
| *Disruption* | 0.800 | | |
| *Support* | 0.727 | | |
| *Parent involvement* | 0.772 | | |
| *DP connectedness* | 0.786 | | |

## Academic Performance

The IRT analysis for academic performance showed that the GPCM (AIC = 66970.06 fit the students' grade data better than the PCM (AIC = 68059.48). Given the superior fit of the GPCM, the estimates of overall DP academic performance were taken from this model[3]. The separation reliability (i.e., the IRT equivalent of Cronbach's alpha coefficient) of these estimates was very good at .876, showing that the outcome variable for the predictive modelling had a very high level of reliability. The item parameter estimates for each of the DP subjects in the dataset are provided in Appendix B. The distribution of the academic performance outcome variable is presented in Figure 1.

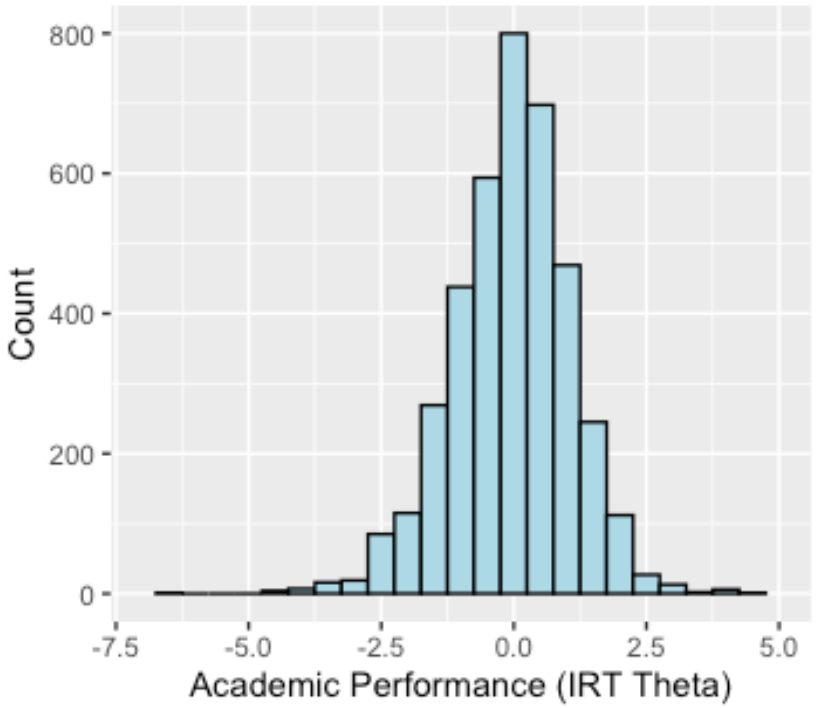


*Figure 1. Distribution of academic performance assessed using IRT scaled DP exam, Extended Essay and Theory of Knowledge grades.*

[3] The correlation between the PCM and GPCM ability estimates was .986 and so this choice had little to no bearing on further findings in the report.

## Wave 1: Beginning of year 1 of the DP

We begin the predictive modelling by examining the Wave 1 predictors of student academic performance using correlations to examine the relationship between each predictor and the IRT scaled academic performance estimates. Given the large number of predictor variables, for each analysis we visualise the 30 items/scales that represent the best predictors of student grades for each academic outcome.

The best predictor of academic performance was the number of hours spent working for pay ($r$ = -.15, $p$ < .001), with students who work more hours tending to perform worse academically. Similarly, the second strongest correlation with academic performance was hour spent doing household chores ($r$ = -.14, $p$ < .001), with students who do more chores tending to perform academically worse. The third strongest correlation, and the strongest positive association, was between hours spent with friends and academic performance, with students who spend more time with friends tending to perform better ($r$ = .13, $p$ < .001). Taken together these results suggest that the way students spend time outside class is an important predictor of academic success. The 30 best Wave 1 predictors (including both the single-item and multi-item scales) of exam grades in terms of correlation coefficients are presented in Figure 2.


*Figure 2. Thirty strongest correlations between Wave 1 variables and academic performance. Error bars are the 95% confidence intervals of the estimates.*

## Wave 2: End of year 1 of the DP

Next, we examine the correlations between Wave 2 variables and academic performance. The variable most strongly correlated with academic performance was hours spent doing homework for DP subjects ($r$ = .15, $p$ < .001), with more hours associated with better performance. The second strongest predictor was bedtime on weeknights, whereby students who go to bed later tended to perform better ($r$ = .15, $p$ < .001), followed by students' self-rated marks compared to their peers in mathematics prior to the DP ($r$ = .14, $p$ < .001). Indeed, many of the other variables that were positively correlated with academic performance were self-rated grades, including self-rated average marks, marks compared to peers in science, and academic abilities. The 30 best Wave 2 predictors (including both the single-item and multi-item scales) of academic performance in terms of correlation coefficients are presented in Figure 3.


*Figure 3. Thirty strongest correlations between Wave 2 variables and academic performance. Error bars are the 95% confidence intervals of the estimates.*

## Wave 3: End of the DP, after the exam session

We turn now to examining the correlation between Wave 3 variables and academic achievement. It is worth noting that self-reported grades, which were some of the strongest positive predictors in Waves 2 were not measured in Wave 3.

The strongest positive correlations with academic performance in Wave 3 were found with the amount of preparation per week a student reported doing for DP exam papers during the exam session ($r = .27$, $p < .001$). The second strongest positive correlation was with the kind of university a student was planning on attending ($r = .20$, $p < .001$), with students indicating they planned on attending a top-level university performing better. The variable with the third strongest relationship with academic performance was indicating that having the Extended Essay as part of one of the other DP subjects would be helpful ($r = -.17$, $p < .001$), with lower preforming students tending to indicate that this would be more helpful compared to higher achieving students. The 30 best Wave 3 predictors (including both the single-item and multi-item scales) of academic performance in terms of correlation coefficients are presented in Figure 4.

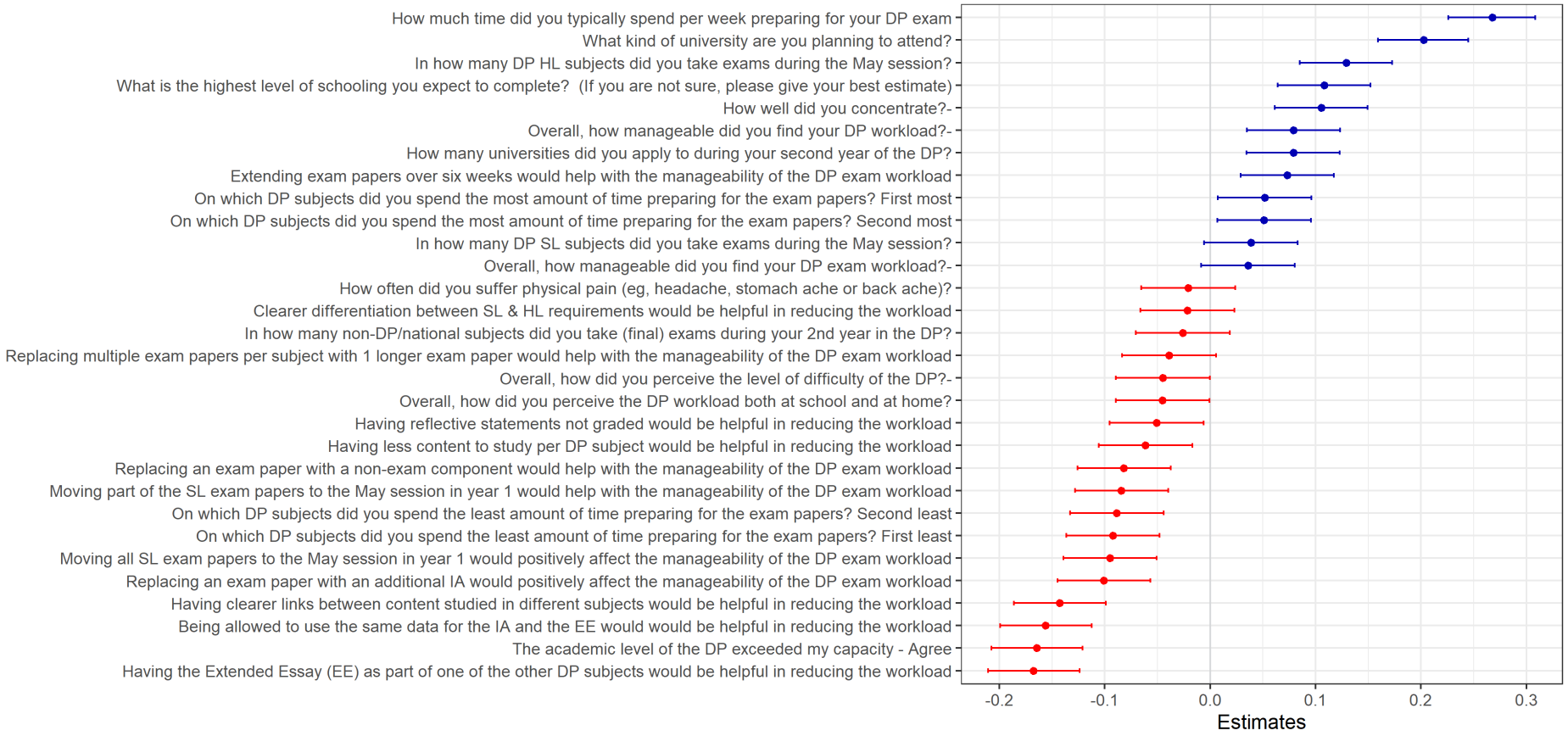


*Figure 4. Thirty strongest correlations between Wave 3 variables and academic performance Error bars are the 95% confidence intervals of the estimates.*

## Variance Decomposition

A variance decomposition model was run to evaluate the percentage of variance in the academic performance measure that occurred between school compared to within schools. Such an analysis is performed by running a multilevel regression model without any predictors, also known as a null model. The model includes random intercepts (i.e., mean performance for each school). From this model, we observed that 42% of the variance in academic performance occurred between schools, while 58% of the variance in academic performance occurred within schools.

## Student Level Models

To examine the student-level predictors of students' academic outcomes, predictors for each wave were modelled using multilevel regression models using the R package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). The final models were fit using stepwise fitting based on the AIC (Chatterjee & Hadi, 2012). The model was fit by comparing the AIC improvements from dropping each candidate variable, and adding each candidate variable from the current model, as well as by dropping or adding the one variable that leads to the best AIC improvement. School was used as the level 2 grouping factor. Compared to correlation, multilevel regression models indicate the unique variance in the outcome explained by a predictor variable, that is, the extent to which a variable predicted the outcome after controlling for all other variables in the model.

### Wave 1: Beginning of year 1 of the DP

The original model estimating academic performance with predictors from Wave 1 included 69 predictors, of which 51 were dropped based on stepwise reduction, leaving a final model with 18 significant predictors. The $R^2$ of the final model was 0.43. The model is presented in Table 2.

The strongest positive predictors of academic performance (after all other covariates were controlled for) were students' self-reported average grades prior to DP, with students who rated their prior grades higher performing better (β= .12). Similarly, students' self-reported grades compared to their peers in science (β= .08) was a significant positive predictor, with students who rated their grades higher performing better. Taken together, these findings suggest that students were able to accurately assess their own academic performance and compare their ability to their peers. The third strongest positive predictor of academic performance was the kind of university the student planned on attending, with students who indicated they planned to attend a top-level university performing better (β= .08).

The model for Wave 1 predictors indicated that the strongest negative Wave 1 predictors of academic performance was whether or not a school was a boarding school[4] (β= -.23) with students who attended a non-boarding school tending to perform worse, and whether a student believed the DP exceeded their capacity (β= -.07), with students who thought the DP

[4] While this variable is obviously related to the school, it was assessed in the student survey and as such was included in the modelling at the school level.

exceeded their capacity tending to perform worse. A visual depiction of the final model organized by the magnitude of the standardized coefficient (beta) is provided in Figure 5. All the effect sizes (betas) would typically be labelled as tiny or small using the conventional cut-offs (Cohen, 1988).

*Table 2. Regression model predicting student academic performance with Wave 1 predictors using stepwise regression.*

| ***Question*** | **β** | **CI (Low)** | **CI (high)** | **p-value** | **Text** |
|---|---|---|---|---|---|
| **A38** | 0.127 | 0.081 | 0.173 | 0.000 | Average grade (mark) before DP |
| **A39c** | 0.088 | 0.042 | 0.134 | 0.000 | Marks compared with classmates in Science before DP |
| **A45** | 0.080 | 0.043 | 0.118 | 0.000 | What kind of university are you planning to attend? |
| **A57c** | 0.078 | 0.041 | 0.115 | 0.000 | My DP teachers grade my homework. |
| **A20a** | 0.062 | 0.025 | 0.100 | 0.001 | How many homework hours do you spend on DP subjects? |
| **A10a** | 0.062 | 0.024 | 0.099 | 0.001 | Number of DP subjects |
| **A22** | 0.058 | 0.021 | 0.095 | 0.002 | Feel about the DP? |
| **A31b** | 0.055 | 0.019 | 0.092 | 0.003 | How many hours do you spend with your (immediate) family? |
| **A84** | 0.050 | 0.015 | 0.084 | 0.005 | Mother: Education level |
| **A32e** | 0.049 | 0.014 | 0.085 | 0.006 | How many hours do you spend reading (e-) books (other than school-related? |
| **A34b** | 0.039 | 0.005 | 0.074 | 0.026 | Sleep per night on weekend days |
| **A30e** | -0.035 | -0.070 | -0.000 | 0.049 | How many hours do you spend doing volunteer work? |
| **A21a** | -0.043 | -0.080 | -0.006 | 0.021 | Additional lessons for DP subjects |
| **A30d** | -0.048 | -0.084 | -0.012 | 0.009 | How many hours do you spend doing household chores? |
| **A67** | -0.052 | -0.089 | -0.015 | 0.006 | Do you have enough energy for everyday life? |
| **A24** | -0.071 | -0.109 | -0.034 | 0.000 | Agreement with: The academic level of the DP exceeds my capacity. |
| **A39a** | -0.087 | -0.130 | -0.044 | 0.000 | Marks compared with classmates in English before DP |
| **A502** | -0.239 | -0.415 | -0.063 | 0.008 | Is your school a boarding school? (No) |

*Note:* All predictors are significant at alpha = .05

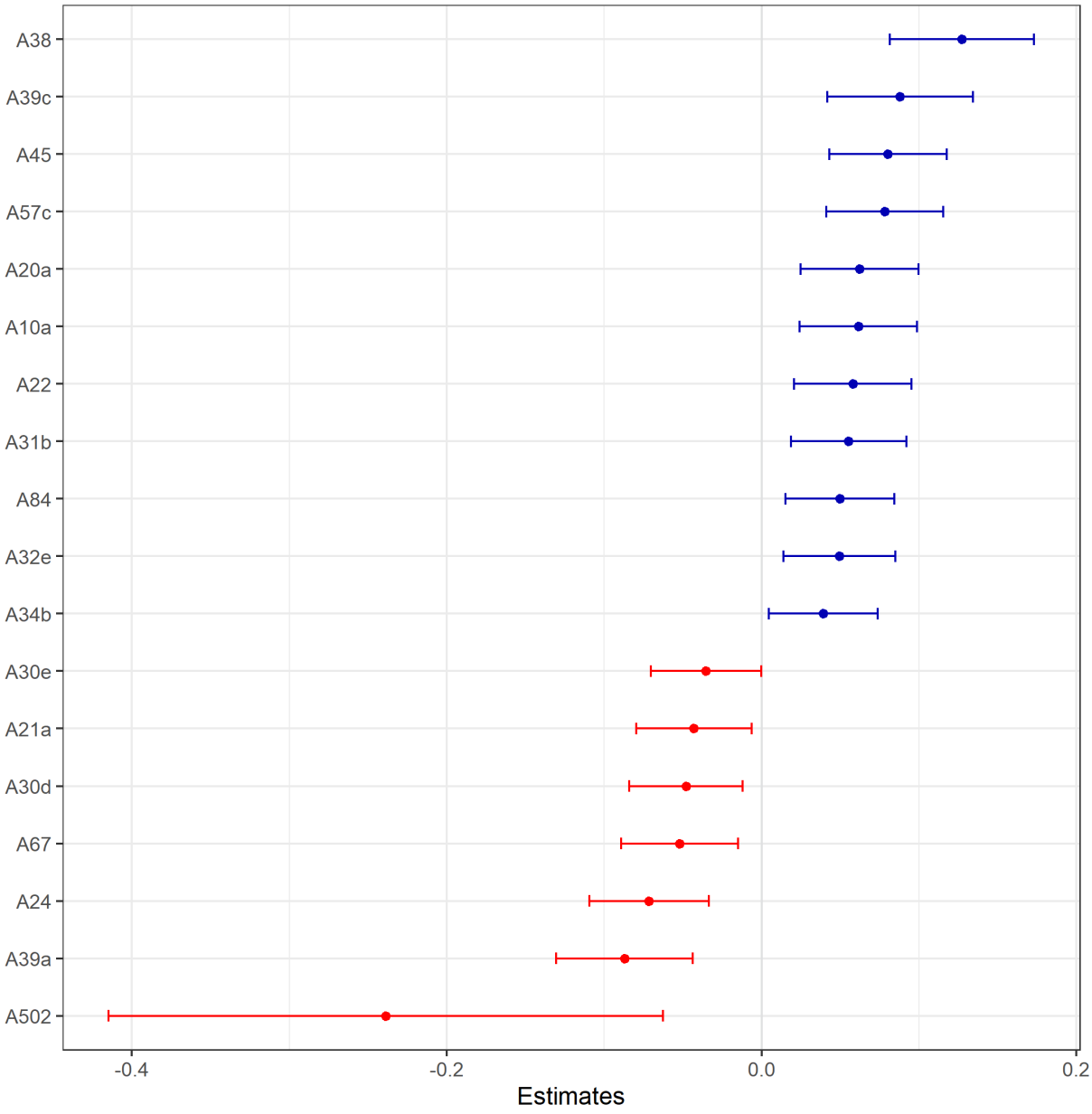


*Figure 5. Standardised regression coefficients for the final model of Wave 1 predictors of academic performance. Error bars are the 95% confidence intervals of the estimates.*

## Wave 2: End of year 1 of the DP

The initial academic performance model based on Wave 2 variables included 69 predictors, of which 53 were dropped based on stepwise reduction, leaving a final model with 16 significant predictors. The $R^2$ of the final model was 0.43. See Table 3 for the model results.

The strongest Wave 2 predictor of academic performance was the year of DP enrolment (β = -.14) with students in their first year performing better than students in the second year. This may be the result of students prioritizing easier exams in the first year of the DP. The second-best predictor of academic performance was hours spent doing homework for DP subjects (β = .10); the more hours students spend on DP homework, the higher their academic performance. All significant effect sizes were again in the tiny to small range (Cohen, 1988) – see Figure 6.

*Table 3. Regression model predicting student academic performance with Wave 2 predictors using stepwise regression.*

| ***Question*** | **β** | **CI (Low)** | **CI (high)** | **p-value** | **Text** |
|---|---|---|---|---|---|
| B10a | 0.099 | 0.063 | 0.135 | 0.000 | How many homework hours do you spend on DP subjects? |
| B26a | 0.076 | 0.038 | 0.115 | 0.000 | How many hours do you spend commuting to and from school? |
| B54 | 0.067 | 0.023 | 0.111 | 0.003 | What year were you born? |
| B36 | 0.066 | 0.027 | 0.105 | 0.001 | How well do you concentrate? |
| B58b | 0.063 | 0.020 | 0.106 | 0.004 | Marks compared with classmates in Mathematics Before DP |
| B55 | 0.060 | 0.024 | 0.096 | 0.001 | What month were you born? |
| B46e | 0.059 | 0.023 | 0.096 | 0.001 | Agreement: Most of my teachers treat me fairly. |
| B57 | 0.045 | 0.002 | 0.087 | 0.041 | Average grade (mark) before DP? |
| B48c | -0.042 | -0.080 | -0.004 | 0.029 | How concerned is your school with wellbeing? |
| B25c | -0.044 | -0.080 | -0.007 | 0.019 | How many hours do you participate in Regular unorganized sports activities? |
| B34 | -0.048 | -0.087 | -0.009 | 0.017 | How well do you sleep? |
| B26c | -0.052 | -0.088 | -0.015 | 0.006 | How many hours do you spend caring for friends or family members who require assistance? |
| B12 | -0.053 | -0.093 | -0.013 | 0.010 | Level of difficulty of the DP? |
| B13 | -0.067 | -0.108 | -0.026 | 0.001 | Agreement: The academic level of the DP exceeds my capacity. |
| B09b | -0.076 | -0.112 | -0.040 | 0.000 | How many hours of class time do you spend on other programme(s)? |
| B062 | -0.141 | -0.220 | -0.062 | 0.000 | Year enrolled |

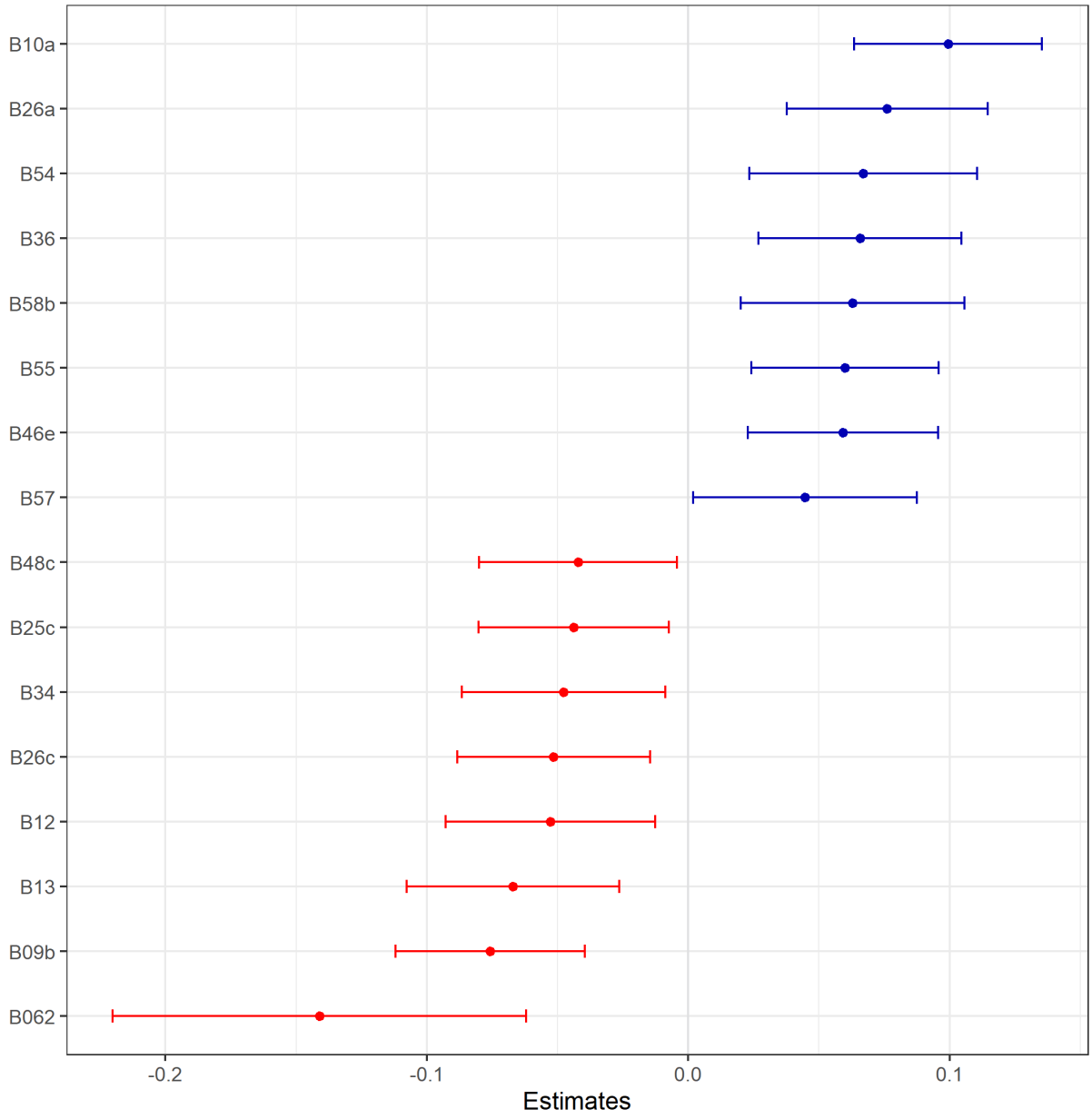


*Figure 6. Standardised regression coefficients for the final model of Wave 2 predictors of academic performance. Error bars are the 95% confidence intervals of the estimates.*

## Wave 3: End of the DP, after the exam session

The initial model based on Wave 3 variable estimating academic achievement included 34 predictors, of which 21 dropped based on stepwise reduction, leaving a final model with 13 significant predictors. The $R^2$ of the final model was 0.45. See Table 4 for model results.

The strongest positive predictor of academic performance was the amount of time spent preparing during the exam session (β = .09), with students who prepared more tending to perform better. The strongest negative predictor of academic performance was the extent to which students agreed with the statement that the DP exceed their capacity (β =-.13), with students who agreed more with the stamen tending to perform worse.

In addition, the Wave 3 predictors of academic achievement indicated the importance of students' post-secondary school plans, including the type of university they planned on attending (β = .09), the number of universities they had applied for (β = .06), and the highest level of schooling they planned on completing (β = .05).

The other factor that seemed pertinent based on the model results was how helpful the students believed different potential changes to the DP would be. Poorer performing students were more inclined to indicate a preference for replacing an exam paper with an additional internal assessment (β = -.06), having clearer links between content studied in different classes (β = -.05) and incorporating the Extended Essay into another DP subject (β = -.06). Higher performing students tended to indicate that having a clearer distinction between higher-level (HL) and standard-level (SL) subjects would be helpful (β = .05), extending the exam period (β = .03), as well as not having reflective statements graded (β = .04). All significant effect sizes were again in the tiny to small range (Cohen, 1988) – see Figure 7.

*Table 4. Regression model predicting student academic performance with Wave 3 predictors using stepwise regression.*

| ***Question*** | **β** | **CI (Low)** | **CI (high)** | **p-value** | **Text** |
|---|---|---|---|---|---|
| C16 | 0.094 | 0.056 | 0.131 | 0.000 | During the exam session, how much time did you typically spend per week preparing? |
| C10 | 0.079 | 0.038 | 0.119 | 0.000 | Overall, how manageable did you find your DP workload? |
| C56 | 0.072 | 0.037 | 0.108 | 0.000 | What kind of university are you planning to attend? |
| C27 | 0.060 | 0.025 | 0.096 | 0.001 | How well did you concentrate? |
| C11a | 0.050 | 0.010 | 0.090 | 0.015 | How helpful do you think having a clearer differentiation between SL and HL would be? |
| C09 | 0.043 | 0.002 | 0.085 | 0.039 | Overall, how did you perceive the DP workload? |
| C11e | 0.043 | 0.007 | 0.079 | 0.021 | Do you think having reflective statements not graded would reduce workload? |
| C21a | 0.034 | 0.000 | 0.067 | 0.050 | Agreement: Extending exam papers over six weeks (instead of the current three-week period) would be helpful |
| C11b | -0.053 | -0.094 | -0.012 | 0.011 | Do you think having clearer links between content studied in different subjects would reduce workload? |
| C21g | -0.057 | -0.091 | -0.023 | 0.001 | Agreement: Replacing an exam paper with an additional IA would be helpful |
| C11c | -0.059 | -0.097 | -0.020 | 0.003 | Do you think having the Extended Essay (EE) as part of one of the other DP subjects would reduce workload? |
| C582 | -0.114 | -0.184 | -0.045 | 0.001 | Did the universities you applied for require you to take an additional test? |
| C07 | -0.134 | -0.174 | -0.095 | 0.000 | Agreement: The academic level of the DP exceeded my capacity |

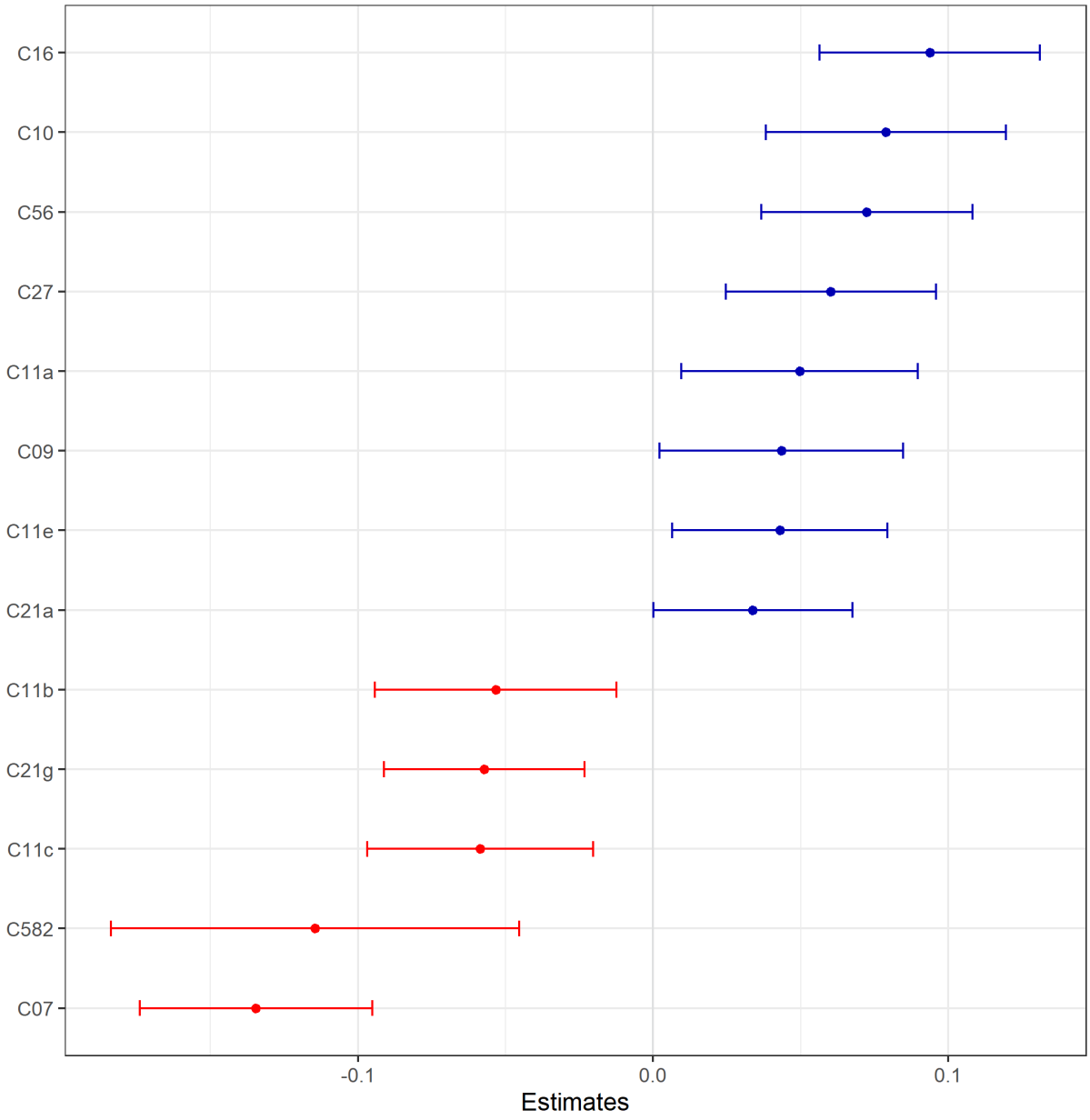


*Figure 7. Standardised regression coefficients for the final model of Wave 3 predictors of academic performance. Error bars are the 95% confidence intervals of the estimates.*

## School Level Models

To examine school-level predictors of academic performance, we again performed a multilevel regression model using stepwise fitting to fit the model based on AIC optimisation. A multi-level regression model was fit with schools as the second-level grouping factor. Based on stepwise reduction, 55 predictors were dropped from the school-level model of academic performance, leaving a final model with 9 significant predictors. The $R^2$ of the final model was 0.44. The model is presented in Table 5.

Factors that positively predicted academic performance were the number of staff members teaching DP subjects (β = .26), with more teachers predicting better performance, as well as the number of students whose first language is different than the language of instruction (β = .18), with schools that have more students whose first language is not the language of instruction tending to do better academically. In addition, schools that report their DP students to be moderately stressed[5] outperformed those that reported students being very stressed (β = -.36). Schools that only sometimes use deadlines for internal assessments tended to perform worse than those that often use them (β = .57) or always use them (β = .49). Schools that used exam results to compare the school with other IB schools tended to outperform those that do not (β = -.25). Finally, schools that implemented *maximum amount of homework per course* outperformed those that do not (β = -.48), while schools who implement *maximum amount of homework per week* tend to perform *worse* academically than those that do not (β = .52). These effect sizes would be considered small-to-moderate in size (Cohen, 1988) – see Figure 8.

[5] No schools reported having DP students with lower than moderate stress and as such the variable only had three utilised response options and was treated as categorical.

*Table 5. Multilevel regression model predicting student academic performance with school-level predictors using stepwise multi-level modelling.*

| ***Question*** | ***β*** | ***CI (Low)*** | ***CI (high)*** | ***p-value*** | ***Text*** |
|---|---|---|---|---|---|
| Q47a4 | 0.568 | 0.185 | 0.950 | 0.004 | To what extent are deadlines for IAs coordinated in your school (Often) |
| Q58d2 | 0.520 | 0.093 | 0.947 | 0.017 | Is maximum amount of homework per week implemented at school? |
| Q47a5 | 0.488 | 0.111 | 0.865 | 0.011 | To what extent are deadlines for IAs coordinated in your school (Always) |
| Q11 | 0.256 | 0.111 | 0.401 | 0.001 | How many of those staff members teach DP subjects? |
| Q15 | 0.176 | 0.067 | 0.284 | 0.002 | First language different to instructional language (% of DP students) |
| Q31a5 | -0.030 | -0.305 | 0.245 | 0.831 | How stressful do you think the DP programme is? (Extremely stressful) |
| Q51h2 | -0.254 | -0.489 | -0.020 | 0.034 | At your school, are DP exam results used to compare the school to other IB schools? |
| Q31a4 | -0.363 | -0.694 | -0.032 | 0.031 | How stressful do you think the DP programme is? (Very stressful) |
| Q58c2 | -0.484 | -0.909 | -0.060 | 0.025 | Is maximum amount of homework per course implemented at school? |

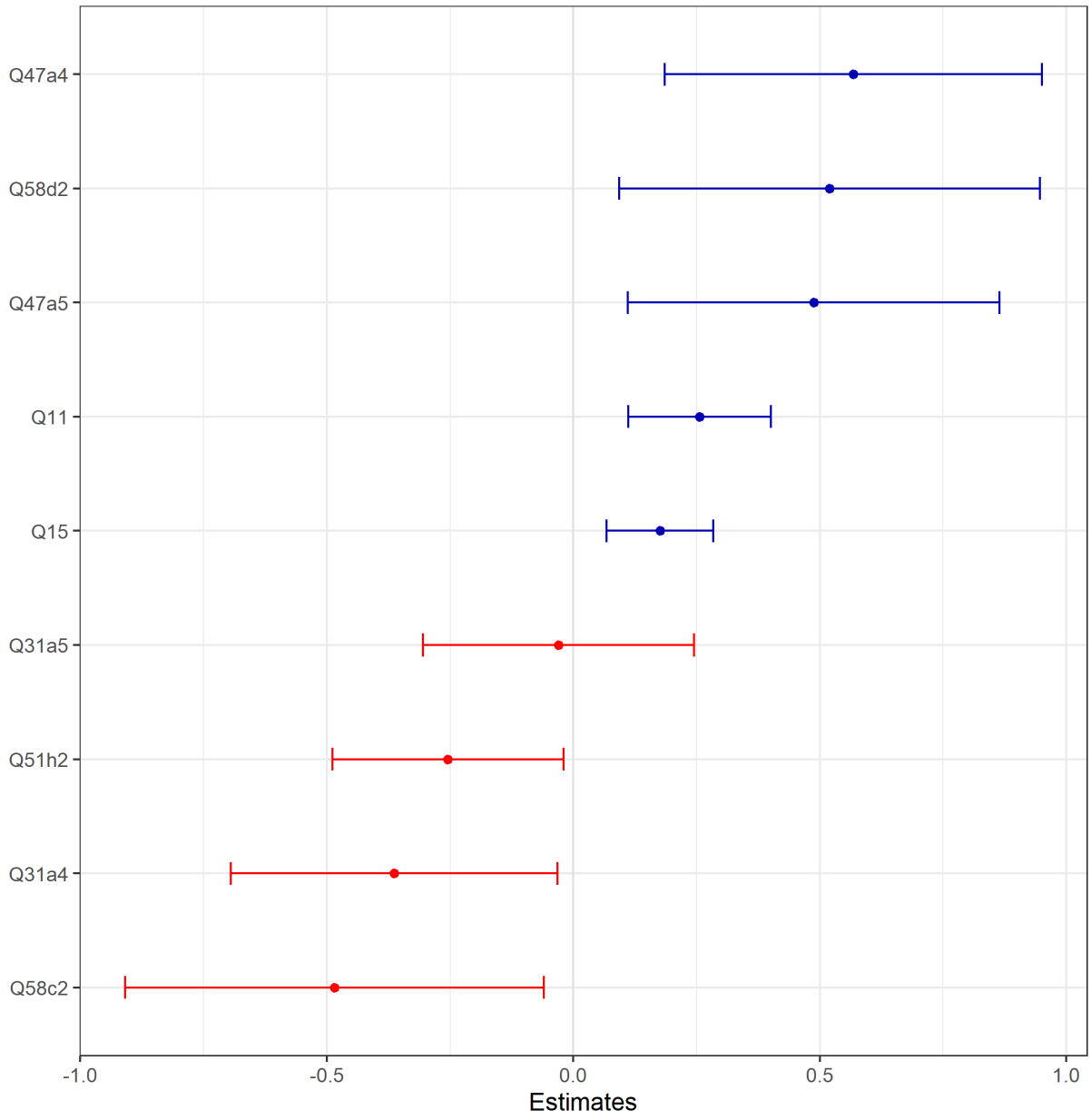


*Figure 8. Standardised regression coefficients for the final model of school-level predictors of academic performance. Error bars are the 95% confidence intervals of the estimates. The following contrasts are applied: (1) moderately stressed vs. very stressed, (2) not at all stressed vs. extremely stressed, (3) never vs often, (5) never vs always.*

## Interactive Models

The aim of the interactive models was to examine the interaction between school-level variables and student-level variables. While it is possible to also consider interactions just at the student or school level, there is a huge pool of potential interactions and there is a risk of model overfitting, and in this case, the sample size was insufficient to consider all possible interactions with a model reduction approach. Even considering just the cross-level interactions, there were many thousands of possible interaction effects, so we therefore opted to take the variables that were significant in the previously run school- and student-level models and calculate cross-level interactions using only these variables. While the interactive models included both main effects and the cross-level School-by-Student interaction effects, we focus our in-text discussion on the prominent interaction effects rather than again discussing the main effects. The full interactive models including main effects are, nonetheless, provided in the subsequent tables.

### Wave 1: Beginning of year 1 of the DP

The interactive wave 1 model, see Table 6, suggested there was a single significant cross-level interaction. Number of DP staff interacted with the number of DP subjects, with the benefit of more staff being more pronounced when there were fewer DP subjects – see Figure 9. This suggests that the benefit of additional staff does not benefit student performance per se, but additional staff time and/or staff per subject predict positive student outcomes. No other interaction effects between significant Wave 1 main effect predictors and significant school-level predictors were significant.

*Table 6. Multilevel model predicting student academic performance with main effects and interaction between significant Wave 1 predictors and school-level predictors using stepwise fitting. Significant interactions are italicised.*

| ***Question*** | **β** | **CI (Low)** | **CI (high)** | **p-value** | **Text** |
|---|---|---|---|---|---|
| Q11 | 0.254 | 0.099 | 0.410 | 0.000 | How many of those staff members teach DP subjects? |
| Q15 | 0.138 | 0.022 | 0.253 | 0.019 | First language different to instructional language (% of DP students) |
| A38 | 0.115 | 0.068 | 0.163 | 0.000 | Average grade (mark) Before DP |
| A39c | 0.092 | 0.044 | 0.141 | 0.000 | Marks compared with classmates in Science Before DP |
| A57c | 0.085 | 0.046 | 0.124 | 0.000 | My DP teachers grade my homework. |
| A45 | 0.075 | 0.036 | 0.115 | 0.000 | What kind of university are you planning to attend? |
| A20a | 0.069 | 0.029 | 0.108 | 0.001 | How many hours do you spend on DP subjects? |
| A22 | 0.064 | 0.025 | 0.103 | 0.001 | Feel about the DP? |
| A32e | 0.060 | 0.023 | 0.097 | 0.001 | How many hours do you spend e. reading (e-) books (other than school-related)? |
| A31b | 0.059 | 0.020 | 0.097 | 0.003 | How many hours do you spend with your (immediate) family? |
| A84 | 0.051 | 0.015 | 0.087 | 0.006 | Mother: education |
| A10a | 0.047 | 0.008 | 0.087 | 0.001 | Number of DP subjects enrolled in |
| A30e | -0.045 | -0.082 | -0.008 | 0.016 | How many hours do you spend doing volunteer work? |
| A21a | -0.046 | -0.084 | -0.007 | 0.019 | Additional lessons for DP subjects |
| *A10a:Q11* | *-0.052* | *-0.101* | *-0.004* | *0.035* | *Number of DP staff by number of DP subjects* |
| A30d | -0.052 | -0.090 | -0.015 | 0.007 | How many hours do you spend doing household chores? |
| A67 | -0.053 | -0.092 | -0.014 | 0.008 | Do you have enough energy for everyday life? |
| A24 | -0.076 | -0.115 | -0.036 | 0.000 | Agreement: The academic level of the DP exceeds my capacity. |
| A39a | -0.081 | -0.126 | -0.035 | 0.000 | Marks compared with classmates in English Before DP |

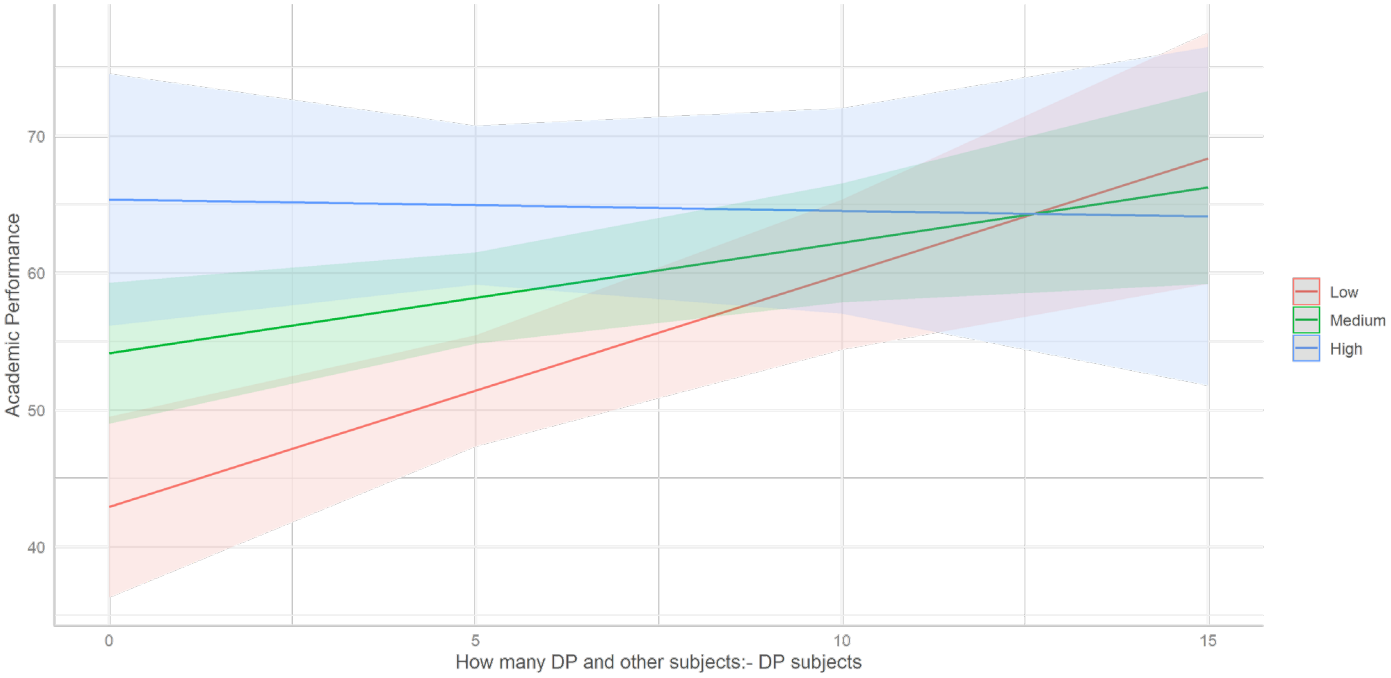


*Figure 9. Interaction between number of DP subjects taken by the students and number of staff members who teach DP subjects (Low, Medium and High). High and low values correspond to ±1SD of the mean, while medium is the mean. Shaded regions = 95% confidence interval*

## Wave 2: End of year 1 of the DP

No significant Wave 2 predictors significantly interacted with any significant school-level predictors, see Table 7.

*Table 7. Multilevel model predicting student academic performance with main effects and interaction between significant Wave 2 predictors and school-level predictors using stepwise fitting.*

| Question | β | CI (Low) | CI (high) | p-value | Text |
|---|---|---|---|---|---|
| Q11 | 0.256 | 0.097 | 0.414 | 0.002 | How many of those staff members teach DP subjects? |
| Q15 | 0.177 | 0.061 | 0.292 | 0.003 | First language different to instructional language (% of DP students) |
| B10a | 0.111 | 0.073 | 0.149 | 0.000 | How many homework hours do you spend on DP subjects? |
| B26a | 0.086 | 0.045 | 0.126 | 0.000 | How many hours do you spend commuting to and from school? |
| B58b | 0.072 | 0.026 | 0.117 | 0.002 | Marks compared with classmates in Mathematics Before DP |
| B54 | 0.068 | 0.022 | 0.114 | 0.004 | What year were you born? |
| B46e | 0.062 | 0.024 | 0.101 | 0.002 | Agreement: Most of my teachers treat me fairly. |
| B36 | 0.058 | 0.017 | 0.099 | 0.006 | How well do you concentrate? |
| B55 | 0.058 | 0.020 | 0.096 | 0.003 | What month were you born? |
| B57 | 0.051 | 0.005 | 0.096 | 0.028 | Average grade (mark) Before DP |
| B48c | -0.043 | -0.083 | -0.003 | 0.036 | How concerned is your school with wellbeing? |
| B25c | -0.048 | -0.087 | -0.009 | 0.015 | How many hours do you participate in regular unorganized sports activities? |
| B34 | -0.048 | -0.090 | -0.007 | 0.023 | How well do you sleep? |
| B26c | -0.055 | -0.094 | -0.016 | 0.006 | How many hours do you spend caring for friends or family members who require assistance? |
| B13 | -0.061 | -0.104 | -0.017 | 0.006 | Agreement: The academic level of the DP exceeds my capacity. |
| B12 | -0.064 | -0.107 | -0.022 | 0.003 | Level of difficulty of the DP |
| B09b | -0.077 | -0.115 | -0.040 | 0.000 | How many class time hours did you spend on other programme(s)? |

## Wave 3: End of the DP, after the exam session

There were two significant cross-level interactions when considering significant Wave 3 predictors. Number of staff members teaching DP subjects interacted with students' ratings of how helpful a clearer distinction between HL and SL subject expectations would be, such that higher performing students were more likely to rate a clearer distinction as helpful when there were more DP staff – see Figure 10. Finally, the positive relationship between exam preparation and academic performance was more pronounced in schools with a high percentage of students whose first language was different from the language of instruction. See Table 8 for model results.

*Table 8. Multilevel model predicting student academic performance with main effects and interaction between significant Wave 3 predictors and school-level predictors using stepwise fitting. Significant interactions are italicised.*

| Question | β | CI (Low) | CI (high) | p-value | Text |
|---|---|---|---|---|---|
| Q11 | 0.213 | 0.056 | 0.370 | 0.268 | How many of those staff members teach DP subjects? |
| Q15 | 0.156 | 0.043 | 0.268 | 0.385 | First language different to instructional language (% of DP students) |
| C16 | 0.101 | 0.061 | 0.141 | 0.514 | During the exam session, how much time did you typically spend per week preparing? |
| C56 | 0.086 | 0.048 | 0.124 | 0.000 | What kind of university are you planning to attend? |
| C10 | 0.083 | 0.040 | 0.126 | 0.000 | Overall, how manageable did you find your DP workload? |
| C27 | 0.066 | 0.028 | 0.105 | 0.001 | How well did you concentrate? |
| C09 | 0.045 | 0.001 | 0.089 | 0.044 | Overall, how did you perceive the DP workload? |
| *Q11:C11a* | *0.045* | *0.007* | *0.083* | *0.019* | *How many of staff members teach DP subjects by How helpful do you think having a clearer differentiation between SL and HL requirements* |
| C11e | 0.044 | 0.005 | 0.083 | 0.028 | How helpful do you think having reflective statements not graded would be? |
| *C16:Q15* | *0.040* | *0.003* | *0.077* | *0.033* | *Time spent preparing by First language different to instructional language (% of DP students)* |
| C21a | 0.038 | 0.002 | 0.074 | 0.041 | Agreement: Extending exam papers over six weeks (instead of the current three-week period) would be helpful |
| C11a | 0.037 | -0.006 | 0.080 | 0.450 | How helpful do you think having a clearer differentiation between SL and HL requirements would be? |
| C11b | -0.048 | -0.092 | -0.004 | 0.031 | How helpful do you think having clearer links between content studied in different subjects would be? |
| C21g | -0.052 | -0.088 | -0.016 | 0.005 | Agreement: Replacing an exam paper with an additional IA would be helpful |
| C11c | -0.056 | -0.097 | -0.015 | 0.007 | How helpful do you think having the Extended Essay (EE) as part of one of the other DP subjects would be? |
| C07 | -0.143 | -0.185 | -0.101 | 0.000 | Agreement: The academic level of the DP exceeds my capacity. |

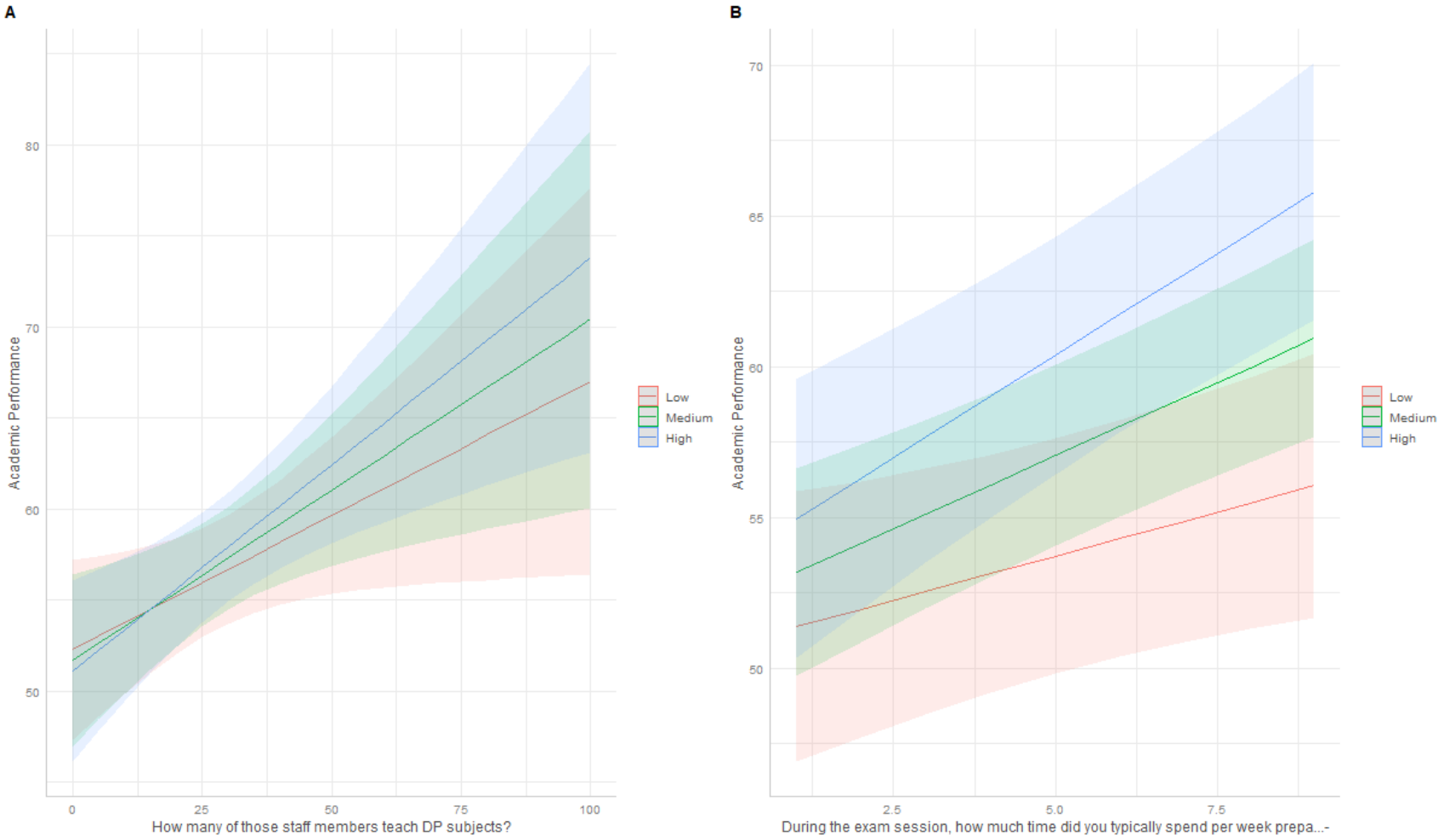


*Figure 10. Interaction between A) number of staff members that teach DP subjects and students' ratings of how helpful a clearer distinction between HL and SL subject expectations would be (Low, Medium and High), and B) time spent preparing for the exam per week during the exam session by percentage of students in the school whose first language was different to the instructional language (Low, Medium and High). Shaded regions = 95% confidence interval.*

## Change Models

Finally, as several scales were repeated across each of the survey time-points, we performed multilevel modelling with the change scores on each of the repeated scales to examine whether these changes predicted academic performance. Wave 1 (beginning of DP Year 1) was viewed as the baseline in the models, with change scores calculated comparing scores in Wave 2 with Wave 1 and comparing scores at Wave 3 with Wave 1. Therefore, positive values on the change score represent an increase in a variable.

The scales repeated across all three time-points were: Happiness, Isolation, Lack of Control, Life Satisfaction, and Time Poorness. Model results are presented in Table 9. The change in happiness from the first to the third survey wave was significant ($\beta$ = .07), with students who became happier across the course of the DP tending to have better academic performance. Conversely, students whose isolation increased from the first to the third survey wave tended to perform worse ($\beta$ = -.05) – see Figure 11.

*Table 9. Multilevel model predicting student academic performance with change scores on time-repeated scales.*

| Scale by Time (T) | β | CI (Low) | CI (high) | p-value |
|---|---|---|---|---|
| Happiness T3 | 0.074 | 0.039 | 0.110 | < .001 |
| Isolation T2 | 0.031 | -0.003 | 0.066 | 0.073 |
| Life Satisfaction T2 | 0.026 | -0.017 | 0.069 | 0.236 |
| Life Satisfaction T3 | 0.015 | -0.025 | 0.054 | 0.465 |
| Time Poorness T2 | -0.004 | -0.044 | 0.036 | 0.837 |
| Time Poorness T3 | -0.006 | -0.044 | 0.032 | 0.745 |
| Happiness T2 | -0.029 | -0.067 | 0.008 | 0.123 |
| Lack of Control T2 | -0.036 | -0.096 | 0.025 | 0.251 |
| Isolation T3 | -0.046 | -0.079 | -0.012 | 0.008 |
| Lack of Control T3 | -0.055 | -0.114 | 0.005 | 0.074 |

*Note: Difference from scores at Wave 1 were used to calculate change scores.*

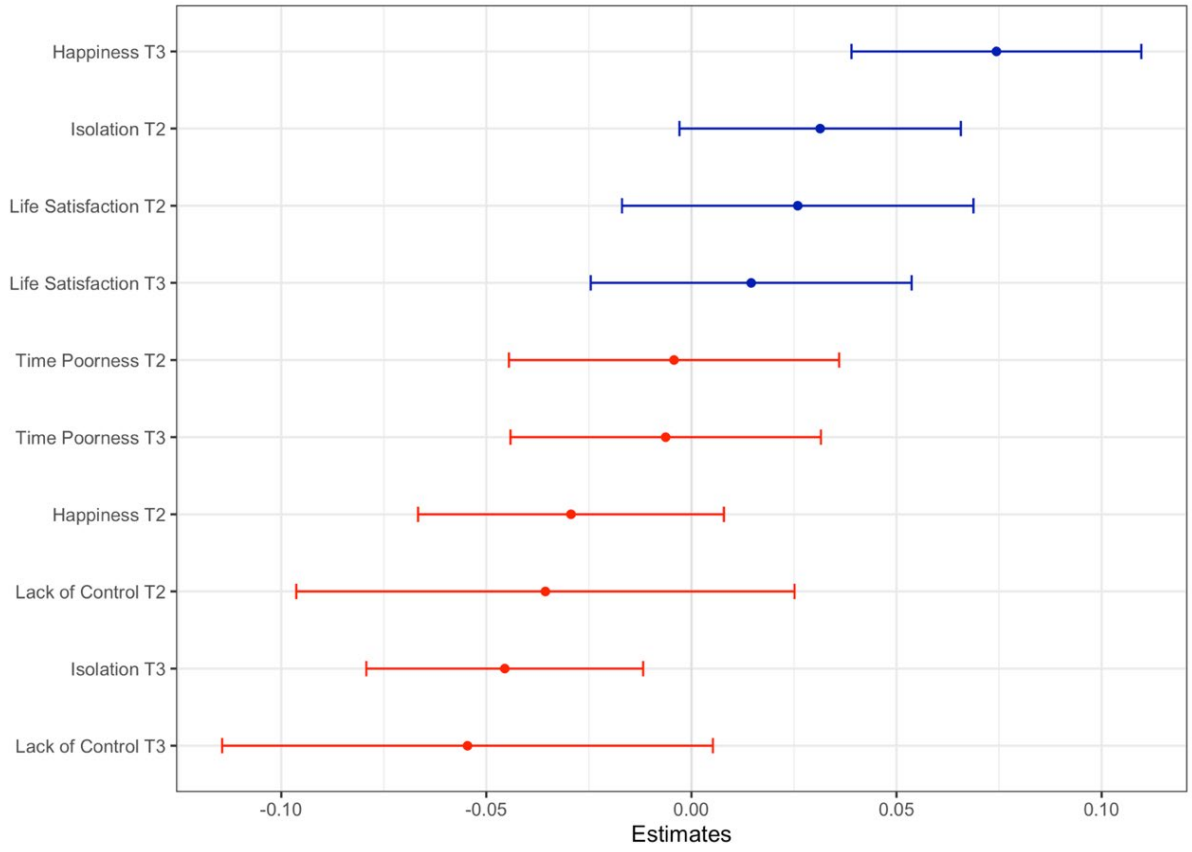


*Figure 11*. Change score predictors of academic performance*. Error bars are the 95% confidence intervals.*

## Discussion

This report outlines a secondary data analysis of student and staff surveys collected over the course of the DP which were used to predict students' academic performance. The results provide insights into the correlates of student academic achievement and the factors that may contribute to academic performance in the Diploma Programme. The report examined bivariate correlations as well as multilevel models to evaluate the relationships between survey variables and academic performance, both in an isolated fashion and with covariates. Overall, the analyses suggested that, among the student-level variables, self-reported grades, students' exam preparation and homework, as well as other activities (e.g., work, caring), were the best predictors of academic performance. While each of the models suggested only a small role for each of the individual predictors, the survey variables were able to collectively account for a substantial portion of variance in academic performance. Furthermore, school-level models showed small-to-moderate effects when predicting student performance using survey responses from DP coordinators. These models indicated the importance of homework and assessment procedures as well as that of students' stress levels. Finally, the change models indicated that improvements in happiness and lower feelings of isolation predicted better academic performance. Below we summarise the main findings of the analyses, with particular focus on the most consistent predictors of academic performance.

Perhaps the most reliable student-level predictors of academic performance were self-reported grades. This is somewhat unsurprising – students generally know their own ability and can accurately predict their performance on past and future academic tasks (Andrade, 2019; Falchikov & Boud, 1989). Indeed, based on a synthesis of meta-analyses, Hattie (2014) found self-rated grades to be the best predictor of academic achievement. While similar findings have sometimes been interpreted causally such that higher self-confidence/self-efficacy is argued to be a pre-cursor to academic achievement (e.g., Stankov, 2013), it remains plausible that the correlation between students' ratings of their abilities and their actual achievement are simply the result of a metacognitive awareness of their own abilities (Jackson, Kleitman, Stankov, & Howie, 2017). Furthermore, we consistently found that poor performing students were likely to indicate agreement with the fact that the academic demands of the DP exceeded their capacity, which supports the stipulation that students are self-aware of their own capabilities.

Variables associated with time spent on learning or preparing for exams were also particularly strong positive predictors of academic performance. As is often the case in the context of high stakes testing, factors associated with learning time that are the most proximal to the assessment task (e.g., exam preparation, time management and practice) are strong predictors of performance (Adesope, Trevisan, & Sundararajan, 2017; Briggs, 2001, Kitsantas et al., 2008). While exam preparation and homework have sometimes been the subject of criticism, as it has been demonstrated that it can narrow the breadth of materials studied by students (e.g., Davis & Vehabovic, 2018), overall, preparing for exams through revision and practice tests is a robust predictor of exam performance (Adesope, Trevisan, & Sundararajan, 2017).

The findings also suggested that involvement in time-demanding activities outside of class may have a negative impact on academic performance. Involvement in paid work and chores were both negatively associated with academic performance. Given the high time-demands of the final years of secondary school, and the DP in particular, it is likely that such activities take time away from study and revision. However, it is worth keeping in mind that such activities may positively benefit students in other ways, and, of course, it is also possible that such variables are indirectly driven by the effect of socio-economic advantage. In support of this socio-economic explanation, other activities such as time spent caring for family or playing unorganised sport negatively predicted performance, while playing organised sport and time spent with family either positively predicted performance or had no relationship with academic performance. These findings suggest that it is not simply time spent revising or studying (or lack thereof) but the way in which time outside of class is spent that predicts academic performance, and this may be due to indirect effects of socio-economic status or home context (Sirin, 2005). Further research is needed to investigate how factors predicting academic performance vary across DP students' socio-economic background, and specifically, several different indicators of socio-economic background at both the student and school levels should be included, as the present study found little to no relationship between parents' education level and students' DP academic performance (Liberatos, Link, & Kelsey, 1988).

In terms of socio-emotional variables, an increase in happiness was associated with better performance, while an increase in isolation was associated with poorer performance. These social-emotional variables have previously been found to be positive predictors of academic achievement (MacCann et al., 2020; Wang, Kiuru, Degol, & Salmela-Aro, 2018). It is also pertinent that the change-from-baseline in these scales predicted performance, suggesting that students who made social and emotional progress throughout the DP were better able to cope with the academic demands of the programme. This may be due to the importance of social groups for coping with stress, particularly during adolescence (Wang, Kiuru, Degol, & Salmela-Aro, 2018).

It is worth noting that the individual effects found in the analyses throughout this report were predominantly only small, yet the overall models predicted large amounts of variance in academic performance. This is somewhat to be expected – academic performance is not the result of any one factor, rather it results from the combination of many factors across the individual student and collective school levels. This suggests that targeting a specific variable through intervention or school policy is unlikely to be successful, rather a holistic approach to improving student learning is more likely to be effective. Importantly, only around half the variance in students' performance[6] was between-schools, suggesting that the contextual factors and policies within schools can play a very important role in driving students' academic performance. This proportion of within school variance observed for the DP schools in the current study is in line with what is typically found when decomposing variance between school and student levels, although this can vary depending on different

[6] It is also worth noting that some of the between-school variance is likely to be due to pre-existing difference in students' ability between schools at the time of enrolment.

school characteristics, such as the enrolment size of the schools, the heterogeneity of their students' background characteristics, and whether they have selective admission or not (Muthén, 1991; Nye, Konstantopoulos, & Hedges, 2004; Leckie & Goldstein, 2019).

The research included in this report is correlational in nature and so one should apply due caution when drawing causal conclusions regarding the relationships between the included student and school variables and DP academic performance. Indeed, while it is often appealing and intuitive to assume students' learning practices like studying, preparing for exams, etc., causally produce better educational outcomes, it cannot be ruled out that such effects are driven by the reverse – that higher achieving students are more likely to engage in such behaviours. In addition, there are further confounds such as third variables, not included in the model, that drive the observed pairwise relationships. For example, a student high in conscientiousness is likely to complete their homework, study more, attend class, and ultimately achieve a better grade, but it becomes extremely difficult to disentangle the effects of each of these practices from one another and from any other unmeasured effects of conscientiousness. As such, the findings presented here need to be supplemented with well controlled experimental studies prior to drawing any strong causal conclusions regarding the determinants of academic performance in the Diploma Programme.

## Recommendations

The current findings suggest that academic performance is driven by the cumulative effect of many student and school characteristics. It is therefore prudent that the IB continue to focus on developing a wide range of student characteristics, especially helping students manage their time and activities, both academic and non-academic, inside and outside of the classroom. While these findings are largely unsurprising, it is comforting to consider that the best preparation for DP examination appears to be the time that students spent inside and outside of class learning, preparing, and revising. The research also indicates that promoting increases in students' happiness and reducing their sense of isolation over the course of the DP may be a promising means of improving their academic performance.

## Conclusions

Overall, the findings presented in this report indicate the role of a diverse set of student and school factors in predicting academic performance in the DP. The findings align well with previous research and suggest that a comprehensive understanding of a multitude of student and school variables, along with their interactions, is necessary when designing educational interventions and policy reforms to improve academic performance. The findings suggest that the strongest predictor of academic performance is students' self-reported academic achievements along with factors related to instructional and learning time, such as class time, study time, and not having to do household chores and work for pay. Moreover, changes in happiness and feelings of isolation throughout the duration of the DP were shown to be associated with final performance. Future research should endeavour to measure more variables over time to allow for more extensive longitudinal analyses, including more indicators of students' home and school context, to develop a temporally richer and well controlled model of predictors of DP academic outcomes.

## References


Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659-701.

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education,* 4, 87.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Ballantyne, K., & Rivera, C. (2014). *Language proficiency for academic achievement in the international baccalaureate diploma programme*. The Hague, Netherlands: International Baccalaureate Organization.

Briggs, D. C. (2001). The effect of admissions test preparation: Evidence from NELS: 88. *Chance*, *14*(1), 10-18.

Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example* (Fifth edition.). Hoboken, New Jersey: Wiley.

Chen, G., & Weikart, L. A. (2008). Student background, school climate, school disorder, and student achievement: An empirical study of New York City's middle schools. *Journal of School Violence*, *7*(4), 3–20.

Cohen, J. (1988). *Statistical power analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates

Davis, D. S., & Vehabovic, N. (2018). The dangers of test preparation: What students learn (and don't learn) about reading comprehension from test-centric literacy instruction. *The Reading Teacher*, 71(5), 579-588.

Fan, X. (2001). Parental involvement and students' academic achievement: A growth modeling analysis. *The Journal of Experimental Education*, *70*(1), 27–61.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*(4), 395-430.

Halic, O. (2013). *Postsecondary educational attainment of IB Diploma Programme candidates from US high schools*. Bethesda, MD: International Baccalaureate Organization.

Hattie, J. (2014). *Visible learning and the science of how we learn*. London, UK: Routledge.

Haynes, N. M., Emmons, C., & Ben-Avie, M. (1997). School climate as a factor in student adjustment and achievement. *Journal of Educational and Psychological Consultation*, *8*(3), 321–329.

Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-regulation and ability predictors of academic success during college: A predictive validity study. Journal of Advanced Academics, 20, 42–68.

Kwok, O. M., Lai, M. H. C., Tong, F., Lara-Alecio, R., Irby, B., Yoon, M., & Yeh, Y. C. (2018). Analyzing complex longitudinal data in educational research: A demonstration with project english language and literacy acquisition (ella) data using xxM. *Frontiers in Psychology*, *9*, 790.

Jackson, S. A., Kleitman, S., Stankov, L., & Howie, P. (2017). Individual differences in decision making depend on cognitive abilities, monitoring and control. *Journal of Behavioral Decision Making*, *30*(2), 209-223.

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(3), 537–554.

Leckie, G., & Goldstein, H. (2019). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal, 45*(3), 518-537.

Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from timss and pisa. *Learning and Individual Differences*, *65*, 50–64.

Liberatos, P., Link, B. G., & Kelsey, J. L. (1988). The measurement of social class in epidemiology. *Epidemiologic Reviews, 10*(1), 87–121.

MacCann, C., Jiang, Y., Brown, L. E., Double, K. S., Bucich, M., & Minbashian, A. (2020). Emotional intelligence predicts academic performance: A meta-analysis. *Psychological Bulletin*, *146*(2), 150.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*(4), 338-354.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237-257.

Pilchen, A., Caspary, K., & Woodworth, K. (2019). *Postsecondary outcomes of IB Diploma Programme graduates in the U.S*. Menlo park, CA: SRI Education.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353.

Robitzsch A, Kiefer T, Wu M (2020). *TAM: Test Analysis Modules*. R package version 3.5-19, https://CRAN.R-project.org/package=TAM.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.

Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences*, *55*(7), 727–732.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357-385.

Thiele, T., Singleton, A., Pope, D., & Stanistreet, D. (2016). Predicting students' academic performance based on school and socio-demographic characteristics. *Studies in Higher Education*, *41*(8), 1424–1446.

Uline, C., & Tschannen-Moran, M. (2008). The walls speak: The interplay of quality facilities, school climate, and student achievement". *Journal of Educational Administration*, *46*(1), 55–73.

Wang, M. T., Kiuru, N., Degol, J. L., & Salmela-Aro, K. (2018). Friends, academic achievement, and school engagement during adolescence: A social network approach to peer influence and selection effects. *Learning and Instruction*, 58, 148-160.

Werblow, J., & Duesbery, L. (2009). The impact of high school size on math achievement and dropout rate. *The High School Journal*, *92*(3), 14–23.

# Appendices

## Appendix A

Correlations between survey instruments:

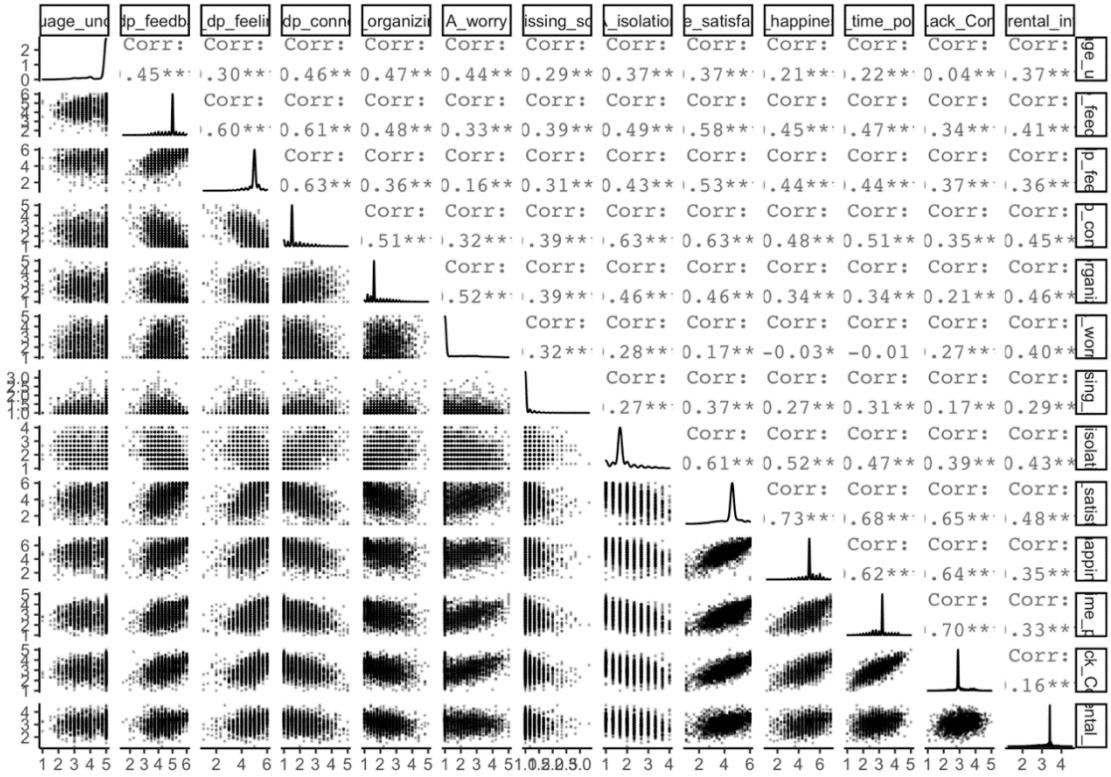

*Figure A-1. Correlations of key multi-item scale items for Survey Wave 1*

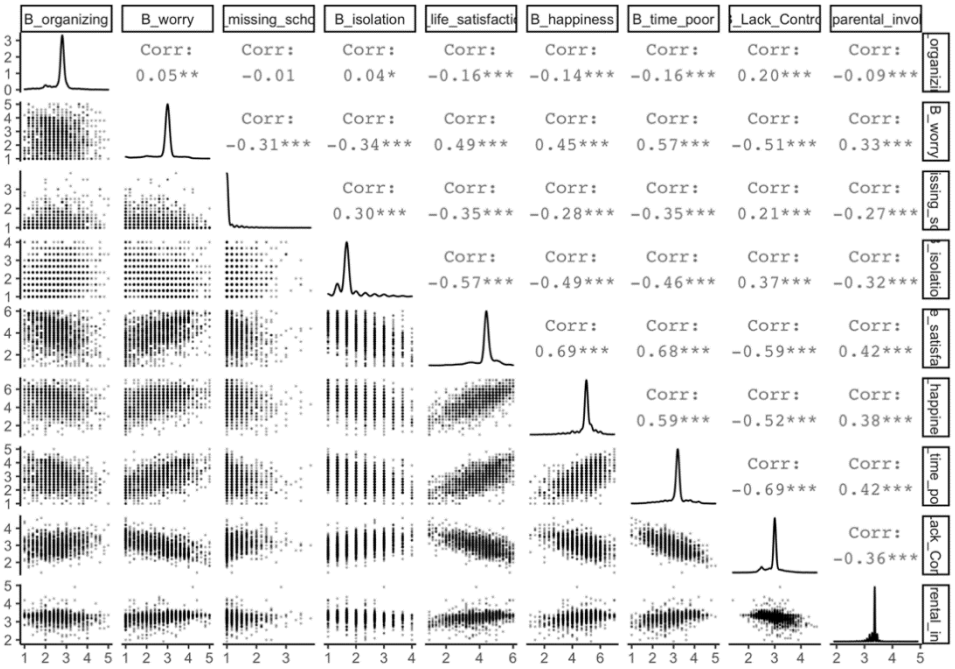


*Figure A-2. Correlations of key multi-item scale items for Survey Wave 2*

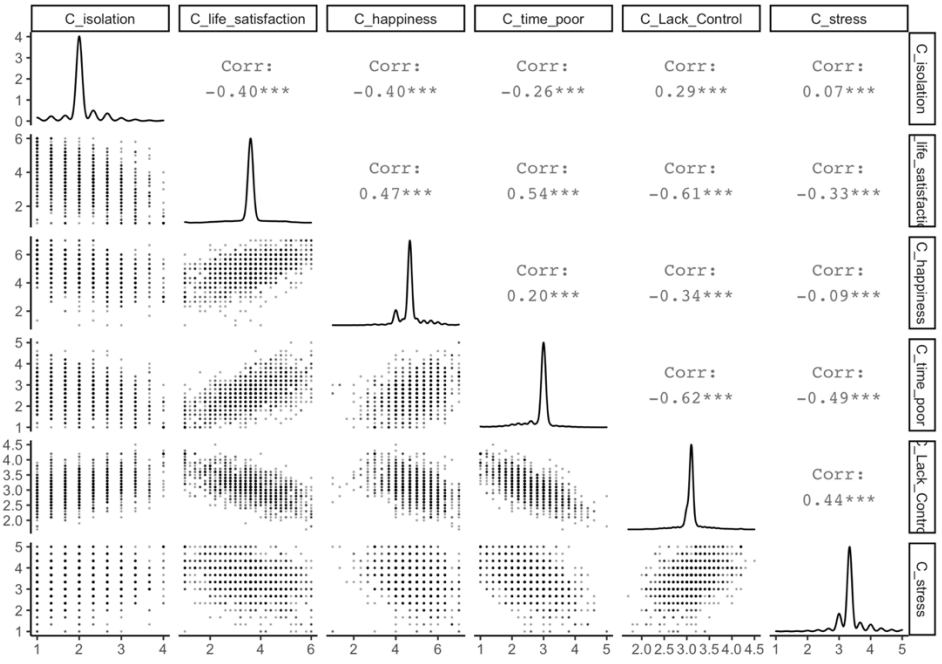


Figure A-3. Correlations of key multi-item scale items for Survey Wave 3

## Appendix B

*Table 10. IRT estimates including the discrimination (alpha), overall item difficulty (beta), and category threshold difficulty (tau) for the different DP subject codes in the dataset.*

| Subject Code | alpha | beta | tau.1 | tau.2 | tau.3 | tau.4 | tau.5 | tau.6 |
|---|---|---|---|---|---|---|---|---|
| 1011 | 1.06 | -0.74 | -2.09 | -1.72 | 0.25 | 1.10 | 2.46 | NA |
| 1012 | 1.60 | -0.68 | -2.17 | -1.37 | -0.10 | 1.29 | 2.35 | NA |
| 1021 | 1.85 | -0.12 | -2.82 | -0.60 | 0.80 | 2.61 | NA | NA |
| 1022 | 1.59 | -0.19 | -2.05 | -0.58 | 0.49 | 2.14 | NA | NA |
| 2011 | 4.75 | -0.12 | -0.52 | 0.13 | 0.39 | NA | NA | NA |
| 2012 | 12.76 | 0.35 | -0.45 | 0.17 | 0.27 | NA | NA | NA |
| 2020 | 0.96 | -1.10 | -2.35 | -2.03 | -0.92 | 0.63 | 1.62 | 3.04 |
| 2031 | 1.12 | -1.06 | -2.58 | -0.95 | 0.05 | 0.98 | 2.50 | NA |
| 2032 | 0.81 | -1.05 | -2.16 | -0.42 | 0.01 | 2.56 | NA | NA |
| 3011 | 3.08 | -0.78 | -1.74 | -0.78 | -0.22 | 0.93 | 1.79 | NA |
| 3012 | 2.66 | -0.98 | -2.16 | -1.69 | -0.53 | 0.66 | 1.35 | 2.37 |
| 3021 | 2.49 | -0.18 | -1.99 | -0.73 | -0.03 | 0.79 | 1.95 | NA |
| 3022 | 3.17 | -0.21 | -1.53 | -0.56 | -0.14 | 0.68 | 1.56 | NA |
| 3031 | 3.40 | -0.39 | -1.34 | -1.65 | -0.05 | 0.79 | 2.26 | NA |
| 3032 | 3.03 | 0.04 | -1.07 | -0.45 | 0.33 | 1.19 | NA | NA |
| 3041 | 2.69 | 0.39 | -1.06 | 0.23 | 0.83 | NA | NA | NA |
| 3042 | 3.92 | 0.48 | -1.00 | -0.11 | 1.11 | NA | NA | NA |
| 3051 | 1.53 | -0.27 | -2.12 | -1.64 | 0.05 | 1.24 | 2.48 | NA |
| 3052 | 2.49 | -0.87 | -3.21 | -1.37 | -0.49 | 0.68 | 1.78 | 2.62 |
| 3061 | 0.93 | 0.04 | -1.20 | -0.69 | 0.23 | 1.66 | NA | NA |
| 3062 | 2.67 | -0.14 | -1.05 | 0.06 | 0.99 | NA | NA | NA |
| 3071 | 1.16 | -0.46 | -2.05 | -0.72 | -0.04 | 2.81 | NA | NA |
| 3072 | 1.39 | -0.43 | -1.05 | -0.07 | 1.11 | NA | NA | NA |
| 3081 | 1.39 | -0.65 | -2.27 | -1.04 | -0.61 | 0.17 | 1.22 | 2.53 |
| 3082 | 2.98 | -0.77 | -2.46 | -1.44 | -0.44 | 0.39 | 1.40 | 2.55 |
| 3091 | 1.60 | -0.25 | -1.18 | -0.60 | 0.39 | 1.40 | NA | NA |
| 3092 | 1.50 | -0.83 | -0.54 | -0.83 | 1.37 | NA | NA | NA |
| 3101 | 0.48 | -1.65 | -2.12 | 2.12 | NA | NA | NA | NA |
| 4011 | 3.39 | -0.66 | -2.53 | -0.97 | -0.20 | 0.47 | 1.23 | 2.00 |
| 4012 | 3.86 | -0.33 | -2.09 | -0.99 | -0.31 | 0.39 | 1.05 | 1.95 |
| 4021 | 3.96 | -0.04 | -2.06 | -0.90 | -0.16 | 0.53 | 0.97 | 1.63 |
| 4022 | 4.02 | 0.14 | -1.78 | -0.92 | -0.18 | 0.39 | 0.92 | 1.57 |
| 4031 | 2.00 | 0.08 | -2.22 | -0.29 | -0.57 | 0.16 | 1.25 | 1.65 |
| 4032 | 2.59 | -0.01 | -1.77 | -1.03 | -0.31 | 0.56 | 0.86 | 1.69 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4041 | 1.51 | -0.67 | -0.27 | -1.96 | 0.30 | 1.93 | NA | NA |
| 4042 | 4.30 | -0.25 | -1.42 | -0.86 | 0.09 | 0.64 | 1.54 | NA |
| 4051 | 2.97 | -0.23 | -2.08 | -0.98 | -0.05 | 0.46 | 1.09 | 1.55 |
| 4052 | 2.82 | 0.31 | -1.51 | -0.51 | 0.15 | 0.68 | 1.19 | NA |
| 4061 | 13.10 | -0.26 | -1.15 | -0.81 | 0.01 | 0.55 | 1.40 | NA |
| 4062 | 2.35 | 0.07 | -0.85 | -0.11 | 0.97 | NA | NA | NA |
| 4071 | 21.52 | 0.43 | -0.74 | -0.39 | 0.05 | 1.08 | NA | NA |
| 5011 | 1.57 | -0.73 | -2.23 | -1.16 | -0.29 | 0.44 | 1.17 | 2.06 |
| 5021 | 2.04 | -0.23 | -2.19 | -1.08 | -0.11 | 0.45 | 1.09 | 1.83 |
| 5022 | 2.44 | 0.03 | -1.89 | -0.65 | -0.29 | 0.34 | 0.98 | 1.52 |
| 5032 | 17.92 | 0.90 | -0.57 | 0.57 | NA | NA | NA | NA |
| 6011 | 0.34 | -1.68 | -4.39 | 4.05 | 0.34 | NA | NA | NA |
| 6012 | 3.95 | -0.76 | -1.04 | -0.73 | 0.69 | 1.07 | NA | NA |
| 6021 | 1.85 | -0.19 | -2.15 | -1.11 | 0.45 | 0.38 | 2.44 | NA |
| 6022 | 2.18 | 0.22 | -1.01 | -0.43 | 0.25 | 1.19 | NA | NA |
| 6031 | 1.45 | -0.86 | -2.38 | -0.26 | 2.28 | 0.36 | NA | NA |
| 6032 | 0.78 | -0.32 | -1.23 | -1.77 | 0.77 | 0.52 | 1.71 | NA |
| 6041 | 1.74 | -0.41 | -1.95 | -1.00 | -1.17 | 0.26 | 1.64 | 2.22 |
| 6042 | 1.22 | -0.74 | -2.58 | -0.69 | -0.86 | 0.23 | 1.01 | 2.89 |
| 6051 | 1.86 | -0.20 | -2.38 | -0.97 | 0.18 | 1.09 | 2.08 | NA |
| 6052 | 1.55 | -0.34 | -1.89 | -0.68 | -0.07 | 0.80 | 1.84 | NA |
| 7021 | 2.78 | -0.09 | -1.69 | -0.55 | -0.02 | 0.79 | 1.47 | NA |
| EE | 0.66 | -0.85 | -4.02 | -0.25 | 1.68 | 2.59 | NA | NA |
| TK | 1.02 | -0.83 | -4.90 | -0.41 | 1.82 | 3.48 | NA | NA |

*Note:* Not all subjects have observations for each category threshold difficulty (tau), as some subjects (EE and TK) only have 5 possible grades, and for some other subjects, either the lower grades and/or higher grades were not observed in the dataset.

## Appendix C

*Table 11. Key for each of the DP subject codes in the dataset including the subject name(s), DP level and number of observations (N).*

| Subject Code | Name(s) | Level | N |
|---|---|---|---|
| 1011 | ARABIC A LIT | SL | 19 |
| | AZERBAI A LIT | | 1 |
| | BOSNIAN A LIT | | 1 |
| | CHINESE A LIT | | 33 |
| | DANISH A LIT | | 1 |
| | DUTCH A LIT | | 3 |
| | ENGLISH A LIT | | 247 |
| | FILIPIN A LIT | | 3 |
| | FINNISH A LIT | | 26 |
| | FRENCH A LIT | | 3 |
| | GERMAN A LIT | | 26 |
| | HEBREW A LIT | | 4 |
| | HINDI A LIT | | 2 |
| | INDONES A LIT | | 1 |
| | ITALIAN A LIT | | 13 |
| | JAPANES A LIT | | 8 |
| | KOREAN A LIT | | 21 |
| | LITHUAN A LIT | | 21 |
| | MACEDON A LIT | | 1 |
| | MALAY A LIT | | 95 |
| | MOD. GR A LIT | | 2 |
| | MONGOLI A LIT | | 1 |
| | NEPALI A LIT | | 1 |
| | NORWEGI A LIT | | 9 |
| | PERSIAN A LIT | | 1 |
| | POLISH A LIT | | 4 |
| | PORTUGU A LIT | | 2 |
| | ROMANIA A LIT | | 1 |
| | RUSSIAN A LIT | | 12 |
| | SESOTHO A LIT | | 6 |
| | SINHALE A LIT | | 1 |
| | SPANISH A LIT | | 74 |
| | SWEDISH A LIT | | 31 |

| | | | |
|---|---|---|---|
| | THAI A LIT | | 3 |
| | TURKISH A LIT | | 1 |
| | URDU A LIT | | 1 |
| | VIETNAM A LIT | | 3 |
| 1012 | CATALAN A LIT | HL | 34 |
| | CHINESE A LIT | | 8 |
| | CZECH A LIT | | 2 |
| | ENGLISH A LIT | | 1293 |
| | FILIPIN A LIT | | 1 |
| | FINNISH A LIT | | 16 |
| | GERMAN A LIT | | 14 |
| | HINDI A LIT | | 1 |
| | ITALIAN A LIT | | 20 |
| | JAPANES A LIT | | 7 |
| | KOREAN A LIT | | 14 |
| | LITHUAN A LIT | | 14 |
| | MALAY A LIT | | 15 |
| | NORWEGI A LIT | | 21 |
| | POLISH A LIT | | 1 |
| | RUSSIAN A LIT | | 3 |
| | SERBIAN A LIT | | 1 |
| | SESOTHO A LIT | | 28 |
| | SLOVENE A LIT | | 24 |
| | SPANISH A LIT | | 200 |
| | SWEDISH A LIT | | 11 |
| 1021 | ARABIC A LAL | SL | 10 |
| | CHINESE A LAL | | 105 |
| | ENGLISH A LAL | | 517 |
| | FRENCH A LAL | | 6 |
| | GERMAN A LAL | | 10 |
| | KOREAN A LAL | | 6 |
| | PORTUGU A LAL | | 3 |
| | SPANISH A LAL | | 20 |
| | SWEDISH A LAL | | 9 |
| | THAI A LAL | | 2 |
| 1022 | ARABIC A LAL | HL | 3 |
| | CHINESE A LAL | | 33 |
| | ENGLISH A LAL | | 927 |
| | FRENCH A LAL | | 14 |
| | GERMAN A LAL | | 14 |

| | | | |
|---|---|---|---|
| | ITALIAN A LAL | | 3 |
| | JAPANES A LAL | | 17 |
| | KOREAN A LAL | | 11 |
| | PORTUGU A LAL | | 6 |
| | SPANISH A LAL | | 101 |
| | SWEDISH A LAL | | 7 |
| | THAI A LAL | | 1 |
| 2011 | LATIN | SL | 21 |
| 2012 | LATIN | HL | 6 |
| 2020 | ARABIC AB. | SL | 7 |
| | ENGLISH AB. | | 19 |
| | FRENCH AB. | | 109 |
| | GERMAN AB. | | 60 |
| | ITALIAN AB. | | 15 |
| | JAPANESE AB. | | 12 |
| | MALAY AB. | | 3 |
| | MANDARIN AB. | | 90 |
| | SPANISH AB. | | 313 |
| 2031 | ARABIC B | SL | 22 |
| | CHINESE B | | 92 |
| | ENGLISH B | | 108 |
| | FINNISH B | | 9 |
| | FRENCH B | | 300 |
| | GERMAN B | | 44 |
| | HEBREW B | | 1 |
| | HINDI B | | 12 |
| | ITALIAN B | | 2 |
| | JAPANESE B | | 4 |
| | KOREAN B | | 1 |
| | NORWEGIAN B | | 3 |
| | PORTUGUESE B | | 3 |
| | RUSSIAN B | | 7 |
| | SPANISH B | | 530 |
| | SWEDISH B | | 17 |
| 2032 | ARABIC B | HL | 66 |
| | CHINESE B | | 36 |
| | DUTCH B | | 1 |
| | ENGLISH B | | 661 |
| | FINNISH B | | 16 |
| | FRENCH B | | 94 |

| | | | |
|---|---|---|---|
| | GERMAN B | | 15 |
| | HINDI B | | 6 |
| | ITALIAN B | | 3 |
| | JAPANESE B | | 2 |
| | KOREAN B | | 1 |
| | NORWEGIAN B | | 4 |
| | PORTUGUESE B | | 4 |
| | RUSSIAN B | | 1 |
| | SPANISH B | | 193 |
| | SWEDISH B | | 34 |
| 3011 | BUS MAN | SL | 217 |
| 3012 | BUS MAN | HL | 514 |
| 3021 | ECONOMICS | SL | 255 |
| 3022 | ECONOMICS | HL | 672 |
| 3031 | GEOGRAPHY | SL | 93 |
| 3032 | GEOGRAPHY | HL | 230 |
| 3041 | GLOB. POL | SL | 16 |
| 3042 | GLOB. POL | HL | 64 |
| 3051 | ART HISTORY | SL | 8 |
| | HISTORY | | 186 |
| 3052 | HISTORY | HL | 1344 |
| 3061 | ITGS | SL | 52 |
| 3062 | ITGS | HL | 38 |
| 3071 | PHILOSOPHY | SL | 83 |
| 3072 | PHILOSOPHY | HL | 74 |
| 3081 | PSYCHOLOGY | SL | 133 |
| 3082 | PSYCHOLOGY | HL | 458 |
| 3091 | SOC.CUL.ANTH. | SL | 42 |
| 3092 | SOC.CUL.ANTH. | HL | 36 |
| 3101 | WORLD RELIG. | SL | 22 |
| 4011 | BIOLOGY | SL | 660 |
| 4012 | BIOLOGY | HL | 1045 |
| 4021 | CHEMISTRY | SL | 459 |
| 4022 | CHEMISTRY | HL | 775 |
| 4031 | COMPUTER SC. | SL | 82 |
| 4032 | COMPUTER SC. | HL | 109 |
| 4041 | DESIGN TECH. | SL | 35 |
| 4042 | DESIGN TECH. | HL | 71 |
| 4051 | PHYSICS | SL | 466 |

| | | | |
|---|---|---|---|
| 4052 | PHYSICS | HL | 478 |
| 4061 | SPORTS EX SCI | SL | 61 |
| 4062 | SPORTS EX SCI | HL | 18 |
| 4071 | ASTRONOMY | SL | 2 |
| | NOS | | 8 |
| 5011 | MATH.STUDIES | SL | 1021 |
| 5021 | MATHEMATICS | SL | 1752 |
| 5022 | MATHEMATICS | HL | 618 |
| 5032 | FURTH. MATHS | HL | 5 |
| 6011 | DANCE | SL | 7 |
| 6012 | DANCE | HL | 16 |
| 6021 | MUSIC | SL | 58 |
| 6022 | MUSIC | HL | 42 |
| 6031 | FILM | SL | 34 |
| 6032 | FILM | HL | 108 |
| 6041 | THEATRE | SL | 84 |
| 6042 | THEATRE | HL | 119 |
| 6051 | VISUAL ARTS | SL | 118 |
| 6052 | VISUAL ARTS | HL | 330 |
| 7021 | ENV. AND SOC. | SL | 320 |
| EE | Extended Essay | | 3352 |
| TK | Theory of Knowledge | | 3332 |