Journal for the Study of Education and Development

What explains Macau students' achievement? An integrative perspective using a machine learning approach

--Manuscript Draft--

Full Title:	What explains Macau students' achievement? An integrative perspective using a machine learning approach
Manuscript Number:	RIYA-2022-0007R2
Order of Authors:	Yi Wang
	Ronnel King
	Joseph Haw
	Shing On Leung
Article Type:	Original Paper
Manuscript Classifications:	Academic Learning & Achievement; Educational Measurement; Educational Psychology
Abstract:	Although Macau students have consistently been recognized as top performers in international assessments, little research has been conducted to explore the various factors that are associated with their achievement. This paper aims to identify factors that could best predict Macau students' reading achievement using PISA 2018 data provided by 2,979 15-year-old students. An integrative theoretical model that considers the critical roles of demographic, personal, and social-contextual factors was used to understand the relative importance of 41 different factors in predicting reading achievement. A machine learning approach, specifically Random Forest Algorithm, was used to analyze the data. Results indicated that variables classified under personal factors (e.g., metacognitive strategies, reading enjoyment, and perceived difficulty) were the most important predictors of Macau students' achievement. A supplementary analysis using Hierarchical Linear Modelling confirmed the findings from the machine learning approach. Implications of the findings are discussed.
Response to Reviewers:	Dear Dr. MÚÑEZ, We are grateful for the opportunity to revise and resubmit manuscript RIYA-2022-0007R1, titled "What explains Macau students' achievement? An integrative perspective using a machine learning approach" to the Journal for the Study of Education and Development. Thank you for your insightful comments that enabled us to improve the quality of our manuscript. We included a point-by-point response to these comments, which is in the attached file "Response to comments". All the comments were properly addressed, and the changes were highlighted in red color in the main text. We summarized the main revisions below: 1.Literature review and the present study In the section on the theoretical framework, we have revised the paragraphs to be more universal and not related to the current study to avoid the interruption of the reading flow. Moreover, we moved and integrated the "Machine learning" section into the methodology part. In addition, we indicated that this study is exploratory in nature in "The Present Study". 2.Results It was suggested that HLM results could not test the robustness of the RF result by the editor. We agreed with this comment. Thus, the conclusion of HLM results was revised. We addressed HLM was used to complement random forest outcomes. 3.Discussion We first discussed the influences of personal factors, which were the most relevant findings in this study.

	4.Whole manuscript We have checked and corrected the grammar and typos of the full text. We believe the comments and feedback have resulted in great improvement to our submission. Please find attached the revision and resubmission. Warm regards, Authors
Funding Information:	

What explains Macau students' achievement? An integrative perspective using a machine learning approach

Abstract

Although Macau students have consistently been recognized as top performers in international assessments, little research has been conducted to explore the various factors that are associated with their achievement. This paper aimed to identify factors that could best predict Macau students' reading achievement using PISA 2018 data provided by 2,979 15-year-old students. An integrative theoretical model that considered the critical roles of demographic, personal, and social-contextual factors was used to understand the relative importance of 41 different factors in predicting reading achievement. A machine learning approach, specifically Random Forest Algorithm, was used to analyze the data. Results indicated that variables classified under personal factors (e.g., metacognitive strategies, reading enjoyment, and perceived difficulty) were the most important predictors of Macau students' achievement. A supplementary analysis using Hierarchical Linear Modelling confirmed the findings from the machine learning approach. Implications of the findings were discussed.

Keywords: reading achievement, PISA, machine learning, Macau education

Introduction

The Programme for International Student Assessment (PISA) is a large-scale assessment aimed at testing 15-year-olds' ability to deal with real-life challenges. The focal subjects for each PISA cycle are reading, mathematics, and science, with reading as the primary focus of the 2018 assessment (OECD, 2019). The PISA results provide evidence of "what works" for educators and policymakers to improve student achievement (Lewis, 2017). One of the trends that have attracted the attention of education researchers and policymakers is that Macau is the only region that has made continuous progress in PISA (OECD, 2019). Since it first participated in PISA in 2006, Macau students' reading performance has risen rapidly from 21st to 3rd place in 2018 (Schleicher, 2019). Macau students' average scores in reading, science, and mathematics ranked 4th in PISA 2015 and 2nd in the most recent wave of PISA in 2018. Hence, it is vital to understand the various factors associated with Macau students' success.

Studies that use PISA have often focused on a limited number of variables (see Karadağ, 2017). Lee and Shute (2010) argued that academic achievement was influenced by a complex host of variables that should be examined in an integrative manner. They proposed the Personal and Social Contextual Factors (PSCF) framework that simultaneously highlighted the role of personal (e.g., student engagement and learning strategies) and social-contextual factors (e.g., school climate and social-family influences), aside from demographic factors, in examining student achievement. Although previous research has found that some factors influenced Macau student academic performance (e.g., grade repetition in Jheng, 2014; socioeconomic status in Ieong et al., 2016), not much is known about the relative contribution of these factors from an integrative perspective.

This study addressed this gap by uncovering key information that may help educational practitioners in sustaining or even facilitating better educational quality. Informed by the PSCF framework, we simultaneously investigated various demographic, personal, and social-contextual variables in the Macau 2018 PISA dataset. We used machine learning to examine which set of variables could best predict academic achievement among Macau students (Yarkoni & Westfall, 2017).

Literature review

Theoretical Framework

Learning is a complex process that is affected by several different factors including affect, behavior, and cognition that occur in a social context (see Bronfenbrenner, 1986; Fraser et al., 1987 for a review). In this regard, learning success is simultaneously determined by multiple factors. To best understand these factors, Lee and Shute's (2010) developed a personal and social contextual framework (PSCF) to identify the most important factors that were associated with academic achievement and compare their relative importance against each other. In their study, constructs influencing students' academic performance could be categorized into personal and contextual factors.

The PSCF framework has the following advantages for exploring key factors of students' achievement. First, it highlighted that students' learning was influenced by several distinct factors working in concert with each other (Lee & Shute, 2010). Second, it was established based on a comprehensive review of factors that impinge on academic achievement. The authors systematically searched and reviewed studies with strong empirical findings in terms of students' academic achievement at the K12 level. Third, it was an extensive and expandable framework that integrated many theories such as social cognition, self-regulation, metacognitive processes, engagement, and other constructs of interest to educators. It was open-ended and relatively broad so that it covered a broad range of factors compared to more specific theoretical frameworks which focus on a limited set of factors. We elaborated on the different types of factors below:

Demographic factors include student background variables such as gender, grade repetition, and grade level (OECD, 2019). Previous studies have examined the relationships between demographic factors and student achievement. For example, gender differences have been found in reading achievement. Several studies have consistently reported that girls perform significantly better than boys on reading tests (Hochweber & Vieluf, 2018). Grade repetition refers to a school practice that requires low-performing students to remain in the same grade level instead of moving up with their peers. Previous studies have identified grade repetition as an important demographic factor affecting academic achievement (e.g., Stankov & Lee, 2017). Finally, other demographic factors such as grade level and immigration status also are significantly associated with reading achievement (Yovanoff et al., 2005). For example, students increased their vocabulary and practice their reading skills over time, so their reading achievement might develop as their grade level increases (Yovanoff et al., 2005).

Personal factors are characteristics within the student, including psychological, cognitive, and behavioral factors. Constructs that fall under the personal dimension include the enjoyment of reading, engagement, and learning strategies. Enjoyment of reading concerns the sense of pleasure derived from reading activities and reflects one's intrinsic motivation (Wigfield, 1997). Reading engagement refers to engagement with reading and can have cognitive, affective, and behavioral components (OECD, 2009a). Learning strategies refer to students' awareness of using effective tactics in the reading process (OECD, 2019). Personal factors are significantly associated with students' reading achievement. For example, Lee and Jonson-Reid (2016) found that academic self-efficacy was positively related to reading achievement. In addition, enjoyment of reading was found to be closely related to reading achievement (Howie et al., 2017).

Psychological well-being is another important personal factor. Previous studies have shown that students who performed better academically also had higher psychological well-being (e.g., Author, 2015). Metacognitive strategy refers to an individual's ability to monitor their performance, develop plans, formulate next steps based on current learning, and regulate their thought processes (Dunlosky & Thiede, 1998). Recent research has found that training students' metacognitive skills could help them understand and monitor their reading processes and outcomes, which can further improve their reading performance (Zhang & Seepho, 2013).

Social-contextual factors refer to the students' social environment. Examples of social-contextual factors include discriminatory school climate (i.e., teachers' stereotypical attitudes toward students from different cultural groups), teachers' stimulation of reading (i.e. supportive instructional practices that frequently encourage students to express their opinions or ask questions to motivate students to actively participate in reading activities), and teacher-directed instruction (i.e., supportive instructional practices in which teachers set clear goals and check whether students have understood the knowledge). Previous studies have frequently examined the effects of these variables on student academic achievement. For example, discriminatory school climate was negatively associated with academic achievement (Thomas et al., 2009), while teachers' supportive reading practices were associated with positive changes in students' reading motivation, engagement, and achievement (Guthrie et al., 2013).

Along these lines, the OECD (2019b) reported that teachers' encouragement of reading was strongly and positively associated with students' enjoyment of reading, which in turn improved their reading achievement. A recent study found that teacher-directed instruction (i.e., teachers explain the idea of learning to students) had both direct and indirect effects on student achievement (Cairns, 2021). In addition, environmental resources have been found to influence student reading achievement. For example, Gubbels et al. (2020) found that students who had moderate access to information and communication technology (ICT) resources had the best reading achievement.

However, past empirical studies of reading literacy have focused only on a few important predictors such as socioeconomic status (SES, Chen & Cui, 2020), psychological well-being (Bücker et al., 2018), or instructional strategies (Guthrie et al., 2013). Hence, not much is known about how these factors are associated with reading achievement. We addressed this gap by employing the integrative PSCF framework in identifying the key predictors of student achievement.

Macau context

Macau is a special administrative region of China. Since colonial times (1557-1999), Macau has been a melting pot of Western and Eastern cultures. Its special history contributes to the development of a unique educational system in Macau, different from that of other regions in Greater China. Under Portuguese rule, the education system was administered using a laissez-faire approach (Tang et al., 2018). As a result, the educational context in Macau is characterized by the prevalence of private schools, the diversity of curricula and teaching, and the absence of a public examination system for graduation (Xie et al., 2018).

Since 1999, the various school systems have been unified into a typical Chinese school system (i.e., 6 years for elementary school, 3 years for middle school, and 3 years for high school) similar to that of Mainland China. Unlike other regions of China where public schools are more widespread, private education is particularly prevalent in Macau. As of the 2020/2021 academic year, 86.0% of the 121 schools in Macau are private schools (DSEJ, 2020). Schools also differ in their language of instruction (Wong, 2019). Although the official languages in Macau are Chinese and Portuguese, Portuguese is not widely spoken and used as a medium of instruction by only 5% of schools (DSEJ, 2020). Chinese (83%) is the most widely used language of instruction in Macau's schools, followed by English (12%) (DSEJ, 2020).

Previous studies have identified some personal and demographic factors, such as self-efficacy and gender, that influence Macau students' performance (Ieong et al., 2016; Mak et al., 2017). For example, Ieong et al. (2016) found that the effects of school SES on students' performance can be mediated by self-efficacy. However, studies on social-contextual factors, e.g., school and family environmental elements, that contribute to academic achievement are also lacking (Gubbels, 2020; Thomas et al., 2009). However, we are not aware of any previous study that has investigated the relative importance of various personal and social contextual variables that were associated with Macau students' achievement.

The present study

This study aimed to identify the most influential predictors of student achievement in Macau using a machine learning approach. Since this study was based on the PISA 2018 assessment, which focused on the reading domain, we focused on students' reading achievement. Informed by the PSCF framework, we used key demographic, personal, and social-contextual factors to understand how these different sets of variables were associated with reading achievement in Macau. The study did not test the PSCF model itself. It was exploratory in nature using the PSCF to house the study, select factors from the PISA dataset, and classify them into different domains. The research framework was shown in Figure 1.

<Insert Figure 1 here>

Methodology

This research drew on the Macau 2018 PISA dataset (https://www.oecd.org/pisa/data/2018database/), which consisted of 3,775 15-year-old students nested in 45 schools across Macau. The contextual questionnaire and the cognitive test were administered in three languages (i.e., Chinese, English, and Portuguese). Due to the special linguistic context of Macau, we limited the sample to participants who used Chinese as the language of the test (79.0%), as the majority of schools adopted Chinese as the test language (n = 2,979; 49% female).

Dependent Variables

We operationalized student reading achievement based on the ten plausible values in reading provided in the PISA dataset. Drawn from the population distribution, these plausible values represented the scores that individuals could reasonably obtain given the complex design of the assessment (OECD, 2009b; Rutkowski et al., 2013). Following the PISA data analysis manual (OECD, 2009b), we analyzed each plausible value in reading and then aggregated them to generate the final estimates. Using these ten plausible values to index reading achievement was considered the gold standard in PISA analysis compared to just using one plausible value (OECD, 2019).

Independent Variables

This study used a combination of background variables and composite variables that represented reading-specific personal and contextual information from the PISA questionnaire. This study initially selected variables that significantly correlated with reading achievement in previous studies. Next, the variables that were used to derive another variable were removed to reduce redundant information. For example, parental education was used as one of the indicators of students' economic, social, and cultural status (ESCS). Hence, we only used ESCS and did not include parental education. Consequently, a total of 41 variables were used in this study. These variables were scaled using the OECD mean with a standard deviation of -1 to +1. For brevity, only the twenty most influential variables found in this study were presented in Table 1. Please refer to the supplementary materials (Table S1) for the full variable list. All the scales had acceptable internal reliabilities, with Cronbach's alpha ranging from 0.66 to 0.90.

Analytic strategy

The Macau PISA 2018 dataset had a low volume of missing data, with only 11 observations having more than 50% missing values. Markov Chain Monte Carlo (MCMC) imputation was conducted via the *mice* package in R to handle the missing data (van Buuren & Groothuis-Oudshorn, 2011). Given the high accuracy rate of MCMC in imputing missing data, all observations were used in the study. The *bifiesurvey* package was used to generate the descriptive statistics and estimate the independent to dependent variable correlation. The preliminary analysis followed the data analysis procedures of PISA (OECD, 2009c).

Machine learning, a type of artificial intelligence that could automatically learn and improved from previous information to accurately predict outcomes, was used in this study. It could detect significant relationships, trends, patterns, exceptions, and anomalies that would otherwise go unnoticed (Sumathi & Sivanandam, 2006). Machine learning approaches have been shown to be applicable to the educational context and could help educators make evidence-based interventions accordingly (Chen et al., 2019; Kiray et al., 2015; Sinclair et al., 2021).

Compared to classical statistical approaches, machine learning methods can recognize the multivariate and complex nature of different predictor variables. First, most classical statistical techniques may encounter overfitting problems, when models are incorrectly fitted to adapt to sample-specific noise (Yarkoni & Westfall, 2017). Machine learning approaches minimize overfitting by using k-fold cross-validation, a subsequent model validation method to determine the optimal number of predictors (e.g., Martínez-Abad, 2019; Martínez-Abad et al., 2020). Moreover, the cross-validation procedure in machine learning streamlines a model by selecting variables that have the greatest contribution in predicting an outcome, thereby increasing the model's performance accuracy. Last, machine learning results are not affected by assumptions (e.g., sample size and collinearity) that strongly influence the *p*-value (Lu & Ishwaran, 2018). In this regard, machine learning can be used as a complementary technique in analyzing large-scale data with a complex combination of variables.

In this study, we first conducted a preliminary analysis that evaluated four different state-of-the-art tree-based ensemble machine learning algorithms (i.e., Random Forest, Gradient and Extreme Gradient Boosting, Extra Tree, and TreeBag) in terms of their accuracy. Results showed that Random Forest regression algorithm performed slightly better than others (See *Supplementary Material II* in the supplementary file). Thus, we used random forest regression algorithm as our primary statistical method.

Random forest takes three-fourths of the sample to train the model and uses the rest of the unused set to cross-validate its prediction. The algorithm iterates the process until the prediction error was minimized or the stopping variable was reached (Yarkoni & Westfall, 2017). The random forest algorithm avoided model-data over-fitting and enhanced prediction accuracy through this cross-validation. Random forest's main metric of model goodness was accuracy rate and its predictive effect. This study used the *ranger* package (Wright et al., 2021), a fast implementation of

Breiman's (2001) random forest algorithm to test these metrics. Sampling weights provided were also considered in running the algorithm. Root Mean Squared Error (RMSE) was reported for prediction accuracy. In general, lower RMSE indicated a higher accuracy rate. The coefficient of determination (R^2) explained the variance of achievement.

The random forest algorithm quantifies the relative importance of each variable in the prediction model. Variable importance was obtained by permutating the values of the variables and computing the decrease in prediction accuracy resulting from this permutation. The variable importance metrics allowed for the selection of the top predictors of reading achievement.

The study conducted a 10-fold cross-validation with five repeats to streamline the regression model and empirically select the strongest predictors of reading achievement. The k-fold cross-validation procedure evaluated the model's prediction performance by sequentially increasing the number of predictors in the model (in order of their importance). We used $k = 10$, as the estimate of prediction error was almost unbiased in a 10-fold cross-validation for selecting models (Simon, 2007). Finally, the study tested a final model using the most important variables with optimal prediction accuracy.

As the random forest's variable importance was a metric of accuracy, it did not provide an estimate that captures the variable's association with reading achievement. Classical statistics such as Hierarchical Linear Model (HLM) can provide this missing information. Hence, we conducted an ancillary HLM analysis to complement the results of the random forest algorithm. As the data was nested in schools, the use of HLM allowed us to control the potential effects of clustering and focus on the individual-level study variables (i.e., fixed effects).

4. Results

Descriptive Statistics and Correlations

A snapshot of descriptive statistics and bivariate correlation of the top 20 predictors can be found in Table 1. Please find the overall results, reliabilities, and description of all variables in the supplementary file (Table S1).

<Insert Table 1 here>

Variable Importance from the Random Forest Algorithm

A random forest regression model consisting of 41 independent variables was tested. The algorithm was set to test 13 variables at each split in the decision tree and was prompted to randomly grow 1000 decision trees. The regression model had a predictive performance of 4366.32 (RMSE=66.28) and could explain 43.50% (R^2=.44) of the variance in reading achievement. Figure 2 showed the predictors ranked in order of their relative importance. Results revealed that the best predictors consist of metacognitive strategies, enjoyment of reading, self-concept, and demographic factors such as grade and grade repetition.

<Insert Figure 2 here>

Figure 3 showed the step-by-step performance of an incremental model created by iterating five times the 10-fold cross-validation procedure. Results showed that the regression model with the top 20 important predictors yielded the lowest prediction error, hence it was the most predictive model.

<Insert Figure 3 here>

Optimal Random Forest Model

We then created and analyzed a final model with only the top 20 predictors. The final model explained 43.20% of the variance with an error rate of 4378.47 (R^2=.43, RMSE=66.17). Figure 4 showed the final ranking of variables by their relative importance in the final model. For example, regarding demographic factors, the most relevant variables in predicting student reading performance were grade repetition (ranked 3rd) and grade (4th). Removing Grade repetition and grade contributed to the model misfit with RMSE values of 42.33 and 35.92, respectively. The ranking of all key factors was presented in Figure 4.

<Insert Figure 4 here>

Supplementary Analysis using HLM

A supplementary HLM was conducted to determine the explanatory effects of each variable in the optimal set of predictors identified by the random forest. HLM results showed that even when school-level effects were accounted for, the topmost variables had statistically significant effects on reading achievement (see Table S2 in the supplementary materials). The effect size of each predictor broadly reflected the importance of the variable found by the random forest algorithm, which complemented our primary results.

5. Discussion

This study aimed to investigate the main factors predicting students' reading achievement in Macau PISA 2018. Using a machine learning approach, we built an explanatory model for reading achievement. We identified 20 key factors that have been frequently examined in previous studies from three categories: demographic (i.e., two variables), personal (i.e., thirteen variables), and social-contextual factors (i.e., five variables) (Agasisti & Longobardi, 2014; Howie et al., 2017; Chen et al., 2019). We ranked these important predictors and found that personal variables were the most critical factors, followed by demographic factors. School and family contexts impacted reading achievement but exhibited the lowest predictive power. Below, we discussed the main variables identified by the random forest algorithm.

Personal factors

In terms of personal factors, we found that source recognition, summarising, understanding of the material, enjoyment of reading, and self-concept (i.e., reading competence and reading difficulty) were the most important personal factors in predicting reading performance. In addition, perception of difficulty in PISA, eudaimonic well-being, and other psychological factors (i.e., general fear of failure, effort in PISA, mastery goal, job expectancy, and competitiveness) were critical for reading performance.

These findings were congruent with previous studies on psychological factors and learning strategies, such as enjoyment (Botes et al., 2021), competitiveness (Author et al., 2012), meta-cognition (Yadava et al., 2018), and perception of difficulty (Von der Embse et al., 2018). For example, our results were consistent with previous studies on factors predicting reading achievement (e.g., Howie et al., 2017), which confirmed that enjoyment of reading was related to reading achievement by increasing frequency and focus on reading. At the same time, enjoyment of reading was associated with other factors that influence reading performance, such as negative emotions (e.g., anxiety), willingness to communicate, and self-perceptions of performance (see Botes et al., 2021 for a review). Enjoyment of reading was also closely related to the construct of intrinsic motivation, and this study confirms the critical importance of enjoyment of reading for learning (Salikin et al., 2017).

Another example of personal factors was metacognitive strategies that enabled students to better understand and monitor the process of reading and further improve their academic performance (Zhang & Seepho, 2013). Consistent with previous studies (Yadava et al., 2018; Muhid et al., 2020), our study found that metacognitive strategies were closely related to reading achievement. In particular, the metacognitive strategy of assessing credibility was the most powerful predictor of achievement. The metacognitive strategy of summarizing information was also important. Assessing credibility evaluates students' skills to identify the valid, updated, accurate, and unbiased text, and summarizing is a strategy to use synthetic exposition to explain the text in a shorter form than the original text (OECD, 2019). Both metacognitive strategies were fundamental to helping students process text, as they would be activated when individuals think about, monitor, and adjust reading activities (OECD, 2019). The results echoed previous machine learning studies, which also identified assessing credibility and summarizing were the most significant factors affecting Chinese students' reading literacy (Kılıç Depren & Depren, 2021; Koyuncu & Fırat, 2020). Therefore, it is important to equip students with metacognitive strategies and improve their ability to effectively use cognitive resources during the reading process.

Surprisingly, the general fear of failure was positively correlated with student achievement in the Macau sample. This finding seemed to contradict previous studies that had linked fear of failure to unfavorable outcomes (Anoita et al., 2020). However, according to Conroy et al. (2007), fear of failure is the tendency to assess anxiety and threat to the situation, which can have either negative or positive consequences. In other words, it can either prevent them from reaching their optimal potential or motivate successful individuals to achieve high levels of performance. Moreover, fear of failure is associated with some defensive strategies that may be more adaptive in Chinese contexts. For example, Hepper et al. (2013) found that Chinese students reported more defensive strategies due to higher levels of fear of failure compared to Western students. In the Zusho et al. (2005) study, although Asian American students tended to show higher levels of fear of failure than their Western peers, fear of failure did not appear to harm their performance. Another study showed that concern about avoiding appearing incompetent before others, which was closely related to fear of failure, was less harmful in Asian collectivist contexts (Author, 2015). Given the conflicting results, further empirical research is needed to confirm the importance of the fear of failure and its effects on reading achievement not only in the Chinese context but also in other cultural contexts.

Demographics

Among all the background information, we identified two important demographic factors that predicted reading literacy, they were grade repetition and grade level. Repeating a grade could impede students' reading literacy (Stankov & Lee, 2017). This study showed that repeating a grade was the most important factor in predicting reading achievement among all demographic factors examined. The results were consistent with previous studies that reported the negative association between grade repetition and academic achievement. In developed countries, most cases of grade repetition are imposed by schools on students who have made poor progress even though they have regularly attended tutoring sessions designed to help them catch up with peers. However, many studies have shown that grade repetition is detrimental to academic achievement. For example, Brophy (2006) found that repeating a grade only temporarily improved student achievement. Repeaters usually participate in the same curriculum and test material but do not acquire the reading ability for future studies. Hence, they continue to lag behind their peers. Similarly, an analysis that included data from 30 countries with a high proportion of repeaters found that students who repeated secondary or elementary school performed worse in reading than non-repeaters (Ikeda & García, 2014). Repeating a grade affects both academic and non-academic outcomes of students. Conversely, it is possible that students repeat a grade because they are struggling with reading, resulting in undesirable reading performance even in repeated grades. Grade repetition was frequently associated with stress, reduced self-efficacy, and impaired peer relationships (Brophy, 2006). Moreover, repeaters tended to report more negative attitudes toward schools than non-repeaters (Ikeda & García, 2014). Last, students with higher grade levels had better vocabulary knowledge and reading fluency, which in turn led to better reading achievement (Yovanoff et al., 2005).

However, it is also possible that this may be a case of reverse causality. Students who have low levels of reading proficiency may be the ones asked to repeat a grade. This is especially true for individuals with learning disabilities, who are more likely to repeat a grade than those who are nondisabled (Broder et al., 1998). Grade repetition can then further hinder students' psychological well-being and learning progress, thus creating a ' vicious cycle '.

In Macau, there was a high rate of grade repetition in secondary schools, exceeding 10% in 2010. However, subsequent years showed a decreasing trend with the repetition rate marking at 5.2% in 2018 (Social Work Bureau, 2019). Previous studies on Macau students' reading achievement reported that a longer duration of class repetition resulted in these students lagging than their peers in the same grade levels (Jheng, 2014).

Social-context

Among the social-contextual factors, discriminatory environment, teacher-directed instruction, stimulation to read, home educational resources, and ICT resources were identified as important factors. The discriminatory environment refers to the stereotypes, prejudice, and discrimination in school. Previous studies have suggested that a discriminatory school climate, often reflected in teacher discrimination against students, could negatively affect students' academic outcomes (Thomas et al., 2009). Similarly, the current study results suggested that a discriminatory school climate reflected in teachers' low expectations of students, negatively associated with students' reading achievement. One possible reason is that students' perceived discrimination is related to adolescents' psychological well-being (Sellers et al., 2003) and antisocial behaviors (Caldwell et al., 2004), which may negatively affect learning outcomes.

Teacher-directed instruction is well-structured instruction with a clear objective, discussion of questions, and a summary of the previous lesson provided by teachers. Our study also found that teacher-directed instruction was an important factor in explaining and improving students' reading literacy, replicating the findings of OECD that teacher-directed instruction was associated with higher achievement in PISA 2015 (Peña-López, 2016). Teachers who provide a well-structured classroom environment can satisfy students' basic psychological needs for skill development (Lee et al., 2020). In other words, these teachers can help students achieve the desired outcomes and avoid the negative ones, which can further improve students' academic achievement. In addition to instructional strategies, consistent with previous studies, this study identified educational and ICT resources at home as important factors in explaining reading achievement. Previous studies also reported that educational resources and ICT availability at home could support students' reading literacy development (Akyüz, et al., 2014; Hu et al., 2018; Skryabin et al., 2015).

Potential interaction among factors

Different factors may interact in influencing students' academic performance. For example, personal factors are also closely associated with social-contextual factors (Lee & Shute, 2010). Indeed, past studies provided evidence for how different variables interacted with each other. For example, school climate might moderate the association between family structure and academic performance (O'Malley et al., 2015). It is hard to include all combinations of interactions among a large number of factors using classical statistics. Even with few predictors, the potential interactions are often too complex to be tested in multiple linear regression (Hong et al., 2020). The random forest algorithm uses decision trees to implicitly incorporate potential interaction into the regression model. Although it does not specify the significant interaction terms, the random forest algorithm makes interactions easier to accommodate than in linear regression modelling (Hastle et al., 2009). This can be taken as a future research direction to find a more effective way to explore the potentially complex interactions among the key variables.

Limitations and implication

Despite its strengths, this study has several limitations. First, because of the cross-sectional nature of the data, causal inferences could not be drawn. Therefore, longitudinal studies and experimental designs are needed to draw causal conclusions on this topic. Second, all data from PISA 2018 on the variables studied were based on students' self-reported questionnaires. Future studies should use more objective measures (e.g., observations, teacher ratings, peer ratings) to improve data quality and avoid bias due to social desirability. Moreover, due to the very small proportion of students who chose Portuguese as the assessment language in PISA 2018 (n = 34, 0.9%), we removed them from the analysis to avoid bias in the results. This study focused only on students who used Chinese as the testing language in PISA. Although Portuguese is one of the official languages in Macau, it is not commonly used as a language of instruction in most schools. On the contrary, Chinese is the most commonly used language in Macau schools (DSEJ, 2020). Nevertheless, we suggest that future studies could examine the important variables for Portuguese-speaking students' achievement with appropriate sample size, and compare them with the results of Chinese-speaking students.

This study has several implications. Given continuously excellent PISA scores, the educational system in Macau merits investigation. In the extant research, however, there are relatively few studies that have investigated Macau students. The majority of research on PISA and reading, in particular, has been conducted in North America and Western Europe (e.g., Berliner 2020; Manu et al., 2021). This study examined in detail the role of different factors in the Macau context. With an integrative framework, the findings allowed us to know the relative importance of various factors in predicting achievement in the Eastern context. For example, many previous studies suggested that girls commonly outperformed boys in reading in the literature (e.g., OECD, 2014; Hochweber & Vieluf, 2018). However, gender was identified to be less relatively important than other key variables in this study. This shows the necessity and significance of comprehensively considering all important variables together to better understand students' performance. The results of our analysis can provide information to policymakers and help them decide which factors to focus limited educational resources on.

Second, this study has enriched the literature by showcasing the complementarity of the machine learning approach and classical statistics in educational research. For education practitioners, the results uncover key information that facilitates understanding and improving student achievement. This research highlights the important roles of personal factors, which are more malleable than contextual factors, e.g., meta-cognition and reading enjoyment, which could be targets of intervention efforts (Reeve & Brown, 1985; Karemaker et al., 2010). Metacognition skills could be enhanced by meaningful, purposeful social interaction, and reflective promoting interventions (Sandi‐Urena et al., 2011). Reading enjoyment could be fostered through school/home intervention (Villiger et al., 2012). Third, this study found several top contextual factors in understanding students' reading achievement, i.e., school climate, teachers' stimulation and direct instruction, and family resources. Compared with personal factors, contextual factors are less important. However, they also play critical roles in the relationship between personal factors and academic achievement (e.g., Ieong et al., 2016). Policymakers should also target these contextual factors to improve students' performance by increasing their personal advantages. For example, improving the school climate can impact students' achievement by increasing their engagement (Konold et al., 2018). Last, this study provides insight into the PSCF framework. It demonstrates the feasibility of personal and social framework, which provides an avenue for scholars and practitioners to understand the complexity of reading achievement. The findings not only specified the factors in each category in the initial framework but also ranked the importance of the different factors.

Conclusion

The current study examined the key factors that contributed to Macau students' academic success in PISA 2018 using a machine learning approach. We identified the 20 most influential factors that were associated with Macau students' success, enriching existing theoretical work by identifying a wide range of factors that are associated with students' academic achievement. This study also has important practical implications for educators and policymakers in Macau and other regions interested in optimizing student learning achievement.

Reference

Author. (2015).

Author. (2012).

Agasisti, T., & Longobardi, S. (2014). Inequality in education: Can Italian disadvantaged students close the gap?. *Journal of Behavioral and Experimental Economics*, *52*, 8-20. doi:10.1016/j.socec.2014.05.002

Akyüz, G. (2014). The effects of student and school factors on mathematics achievement in TIMSS 2011. *Education and Science, 39*(172), 150-162.

Anoita, B. S., Tiatri, S., & Sari, M. P. (2020). The relationship between fear of failure and student academic achievement with procrastination as mediating variable. In *The 2nd Tarumanagara International Conference on the Applications of Social Sciences and Humanities* (TICASH 2020) (pp. 538-543). Atlantis Press.

Botes, E., Dewaele, J. M., & Greiff, S. (2021). Taking stock: An overview of the literature and a preliminary meta-analysis of foreign language enjoyment and other individual difference variables. *PsyArXiv*. Epub ahead of print. doi:10.31234/osf.io/guaj5.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199-215. doi:10.1214/ss/1009213726

Broder, H. L., Richman, L. C., & Matheson, P. B. (1998). Learning disability, school achievement, and grade retention among children with cleft: a two-center study. *The Cleft palate-craniofacial journal, 35*(2), 127-131.

Brophy, J. (2006). *Grade repetition: Education policy series 6.* International Institute for Educational Planning.

Bücker, S., Nuraydin, S., Simonsmeier, B. A., Schneider, M., & Luhmann, M. (2018). Subjective well-being and academic achievement: A meta-analysis. *Journal of*

Research in Personality, 74, 83-94. doi:10.1016/j.jrp.2018.02.007

Caldwell, C. H., Kohn-Wood, L. P., Schmeelk-Cone, K. H., Chavous, T. M., & Zimmerman, M. A. (2004). Racial discrimination and racial identity as risk or protective factors for violent behaviors in African American young adults. *American Journal of Community Psychology*, *33*(1), 91-105.

Chen, F., & Cui, Y. (2020). Investigating the relation of perceived teacher unfairness to science achievement by hierarchical linear modeling in 52 countries and economies. *Educational Psychology, 40*(3), 273-295. doi:10.1080/01443410.2019.1652248

Chen, J., Zhang, Y., Wei, Y., & Hu, J. (2019). Discrimination of the contextual features of top performers in scientific literacy using a machine learning approach. *Research in Science Education*, 1-30. doi:10.1007/s11165-019-9835-y

Conroy, D. E., Kaye, M. P., & Fifer, A. M. (2007). Cognitive links between fear of failure and perfectionism. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 25*(4), 237-253. doi:10.1007/s10942-007-0052-7

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98*(1), 37-56. doi:10.1016/S0001-6918(97)00051-6

Education and Youth Affairs Bureau (DSEJ). (2020). *Educational statistics*. Retrieved March 17, 2021, from https://portal.dsedj.gov.mo/webdsejspace/internet/Inter_main_page.jsp?id=8525

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational*

Research, 11(2), 147-252.

Gubbels, J., Swart, N. M., & Groen, M. A. (2020). Everything in moderation: ICT and reading performance of Dutch 15-year-olds. *Large-scale Assessments in Education, 8*(1), 1-17. doi:10.1186/s40536-020-0079-0

Guthrie, J. T., Klauda, S. L., & Ho, A. N. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly, 48*(1), 9-26. doi:10.1002/rrq.035

Hepper, E. G., Sedikides, C., & Cai, H. (2013). Self-enhancement and self-protection strategies in China: Cultural expressions of a fundamental human motive. *Journal of Cross-Cultural Psychology, 44*, 5-23. doi:10.1177/0022022111428515

Hochweber, J., & Vieluf, S. (2018). Gender differences in reading achievement and enjoyment of reading: The role of perceived teaching quality. *The Journal of Educational Research, 111*(3), 268-283. doi:10.1080/00220671.2016.1253536

Hong, M., Jacobucci, R., & Lubke, G. (2020). Deductive data mining. *Psychological Methods, 25*(6), 691-707. doi:10.1037/met0000252

Howie, S. J., Combrinck, C., Tshele, M., Roux, K., McLeod Palane, N., & Mokoena, G. (2017). *PIRLS 2016: South African highlights report*. Centre for Evaluation and Assessment (CEA).

Hu, X., Gong, Y., Lai, C., & Leung, F. K. (2018). The relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A multilevel analysis. *Computers and Education, 125*, 1-13. doi:10.1016/j.compedu.2018.05.021

Ieong, M.K., Cheung, K.C., Sit, P.S., Soh, K.C., & Mak, S.K. (2016). Effects of economic, social and cultural status on mathematics performance: A multilevel mediation analysis of self-regulated learning processes. *Contemporary Educational Research Quarterly, 24*(4), 109-143. doi:10.6151/CERQ.2016.2404.05

Ikeda, M., & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies*, *2013*(1), 269-315. doi:10.1787/19952856

Jheng, Y. J. (2014). Does grade repetition work? Who repeats grades? Evidence from the scores of reading literacy of PISA 2009 of Macau. *Jiaoyu Yanjiu Yuekan= Journal of Education Research*, *242*, 97-111. doi:10.3966/168063602014060242007

Karadağ, E. (Ed.). (2017). *The factors effecting student achievement: Meta-analysis of empirical studies.* Springer.

Karemaker, A., Pitchford, N. J., & O'Malley, C. (2010). Enhanced recognition of written words and enjoyment of reading in struggling beginner readers through whole-word multimedia software. *Computers & Education, 54*(1), 199-208.

Kılıç Depren, S., & Depren, Ö. (2021). Cross-Cultural Comparisons of the Factors Influencing the High Reading Achievement in Turkey and China: Evidence from PISA 2018. *The Asia-Pacific Education Researcher,* 1-11. doi:10.1007/s40299-021-00584-8

Kiray, S. A., Gok, B., & Bozkir, A. S. (2015). Identifying the factors affecting science and mathematics achievement using data mining methods. *Journal of Education in Science Environment and Health, 1*(1), 28-48.

Konold, T., Cornell, D., Jia, Y., & Malone, M. (2018). School climate, student engagement, and academic achievement: A latent variable, multilevel multi-informant examination. *Aera Open, 4*(4), 2332858418815661.

Koyuncu, İ., & Fırat, T. (2020). Investigating reading literacy in PISA 2018 assessment. *International Electronic Journal of Elementary Education, 13*(2), 263-275.

Lee, J., & Shute, V. J. (2010). Personal and social-contextual factors in K-12 academic performance: An integrative perspective on student learning. *Educational Psychologist, 45*(3), 185-202. doi:10.1080/00461520.2010.493471

Lee, Y. S., & Jonson-Reid, M. (2016). The role of self-efficacy in reading achievement of young children in urban schools. *Child and Adolescent Social Work Journal, 33*(1), 79-89. doi:10.1007/s10560-015-0404-6

Lewis, S. (2017). Governing schooling through ‘what works’: The OECD’s PISA for schools. *Journal of Education Policy, 32*(3), 281-302. doi:10.1080/02680939.2016.1252855

Lu, M., & Ishwaran, H. (2018). A prediction-based alternative to P values in regression models. *The Journal of Thoracic and Cardiovasccular Surgery, 155*, 1130-1136. doi:10.1016/j.jtcvs.2017.08.056

Mak, S. K., Cheung, K. C., Soh, K., Sit, P. S., & Ieong, M. K. (2017). An examination of student-and across-level mediation mechanisms accounting for gender differences in reading performance: A multilevel analysis of reading engagement. *Educational Psychology*, *37*(10), 1206-1221. doi:10.1080/01443410.2016.1242712

Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M. J. (2020). Educational data mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, *66*, 100875.

Muhid, A., Amalia, E. R., Hilaliyah, H., Budiana, N., & Wajdi, M. B. N. (2020). The effect of metacognitive strategies implementation on students' reading comprehension achievement. *International Journal of Instruction, 13*(2), 847-862. doi:10.29333/iji.2020.13257a

OECD (2009a). *PISA 2009 reading framework.* OECD Publications.

OECD. (2009b). *PISA data analysis manual SPSS second edition.* OECD Publishing. doi:10.1787/9789264056275-en

OECD. (2009c). *Analyses with plausible values. In PISA data analysis manual: SPSS second edition* (pp. 117-131). OECD Publishing. doi:10.1787/9789264056275-9-en

OECD. (2019). *PISA 2018 Assessment and analytical framework.* OECD Publishing. doi:10.1787/b25efab8-en

O'Malley, M., Voight, A., Renshaw, T. L., & Eklund, K. (2015). School climate, family structure, and academic achievement: a study of moderation effects. *School Psychology Quarterly, 30*(1), 142-157. doi: 10.1037/spq0000076

Peña-López, I. (2016). *PISA 2015 results (volume ii): Policies and practices for successful schools.* OECD Publishing.

Reeve, R. A., & Brown, A. L. (1985). Metacognition reconsidered: Implications for intervention research. *Journal of Abnormal Child Psychology, 13*(3), 343-356.

Rutkowski, L., von Davier, M., & Rutkowski, D. (2013). *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis.* CRC Press.

Salikin, H., Bin-Tahir, S. Z., Kusumaningputri, R., & Yuliandari, D. P. (2017). The Indonesian EFL learners' motivation in reading. *English Language Teaching, 10*(5), 81-90. doi:10.5539/elt.v10n5p81

Sandi□Urena, S., Cooper, M. M., & Stevens, R. H. (2011). Enhancement of metacognition use and awareness by means of a collaborative intervention. *International journal of science education, 33*(3), 323-340.

Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. OECD Publishing.

Sellers, R. M., Caldwell, C. H., Schmeelk-Cone, K. H., & Zimmerman, M. A. (2003). Racial identity, racial discrimination, perceived stress, and psychological distress among African American young adults. *Journal of Health and Social behavior*, 302-317.

Simon, R. (2007). Resampling strategies for model assessment and selection. *In Fundamentals of data mining in genomics and proteomics* (pp. 173-186). Springer.

Sinclair, J., Jang, E. E., & Rudzicz, F. (2021). Using machine learning to predict children's reading comprehension from linguistic features extracted from speech and writing. *Journal of Educational Psychology, 113*(6), 1088-1106. doi:10.1037/edu0000658

Skryabin, M., Zhang, J., Liu, L., & Zhang, D. (2015). How the ICT development level and usage influence student achievement in reading, mathematics, and science. *Computers & Education, 85*, 49-58. doi:10.1016/j.compedu.2015.02.004

Social Work Bureau (2019). *Database of Macau children: Grade repetition rate.* Retrieved from https://www.childrendb.ias.gov.mo/www/ratio/search?ratioId=229bd78219f04fd89b74a5a937caadb8

Stankov, L., & Lee, J. (2017). Self-beliefs: Strong correlates of mathematics achievement and intelligence. *Intelligence, 61*, 11-16. doi:10.1016/j.intell.2016.12.001

Sumathi, S., & Sivanandam, S. N. (2006). Introduction to data mining principles.

Introduction to data mining and its applications, 1-20.

Tang, S. Y., Wong, P. M., Wong, A. K., & Cheng, M. M. (2018). What attracts young people to become teachers? A comparative study of pre-service student teachers' motivation to become teachers in Hong Kong and Macau. *Asia Pacific Education Review, 19*(3), 433-444. doi:10.1007/s12564-018-9541-x

Thomas, O. N., Caldwell, C. H., Faison, N., & Jackson, J. S. (2009). Promoting academic achievement: The role of racial identity in buffering perceptions of teacher discrimination on academic achievement among African American and Caribbean Black adolescents. *Journal of Educational Psychology*, *101*(2), 420-431.

van Buuren, S., & Groothuis-Oudshorn, K. (2011). MICE: Multivariate imputation by chained equation. *Journal of Statistical Software, 45*, 1-67. doi:10.18637/jss.v045.i03.

Villiger, C., Niggli, A., Wandeler, C., & Kutzelmann, S. (2012). Does family make a difference? Mid-term effects of a school/home-based intervention program to enhance reading motivation. *Learning and Instruction, 22*(2), 79-91.

Von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, *227*, 483-493. doi:10.1016/j.jad.2017.11.048

Walberg, H. J. (1981). A psychological theory of educational productivity. In F. H. Farley & N. Gordon (Eds.), *Psychology and education* (pp. 81-110). McCutchan.

Wigfield, A. (1997). Reading motivation: A domain-specific approach to motivation. *Educational Psychologist, 32*(2), 59-68. doi:10.1207/s15326985ep3202_1.

Wong, M. (2019). Teaching reading in the Macau secondary English classroom: Some critical issues to consider. In B. Reynolds, M. Teng (Eds.). *English literacy instruction for Chinese speakers* (pp. 195-208). Palgrave Macmillan. doi:10.1007/978-981-13-6653-6_12

Wright, M. N., Wager, S., & Probst, P. (2021). *Ranger: A fast implementation of random forests (0.13.1).* The Comprehensive R Archive Network. doi:10.1080/10618600.2014.983641

Xie, H., Ng, W. S., Zou, D., & Wang, F. L. (2018). A comparative study on recent educational policy changes of primary and secondary schooling in Hong Kong and Macau. In *2018 International symposium on educational technology (ISET)* (pp. 117-120). IEEE. doi:10.1109/ISET.2018.00034

Yadava, S., & Yadava, A. (2018). Cognitive predictors of academic achievement in middle school students. *Indian Journal of Health and Wellbeing*, *9*(1), 158-162.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in Psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. doi:10.1177/1745691617693393

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, *24*(3), 4-12.

Zhang, L., & Seepho, S. (2013). Metacognitive strategy use and academic reading achievement: Insights from a Chinese context. *Electronic Journal of Foreign Language Teaching, 10*(1), 1-16.

Zusho, A., Pintrich, P. R., & Cortina, K. S. (2005). Motives, goals, and adaptive patterns of performance in Asian American and Anglo American students.

Learning and Individual Differences, 15(2), 141-158. doi:10.1016/j.lindif.2004.11.003

Figure 1

Conceptual framework of factors influencing reading achievement

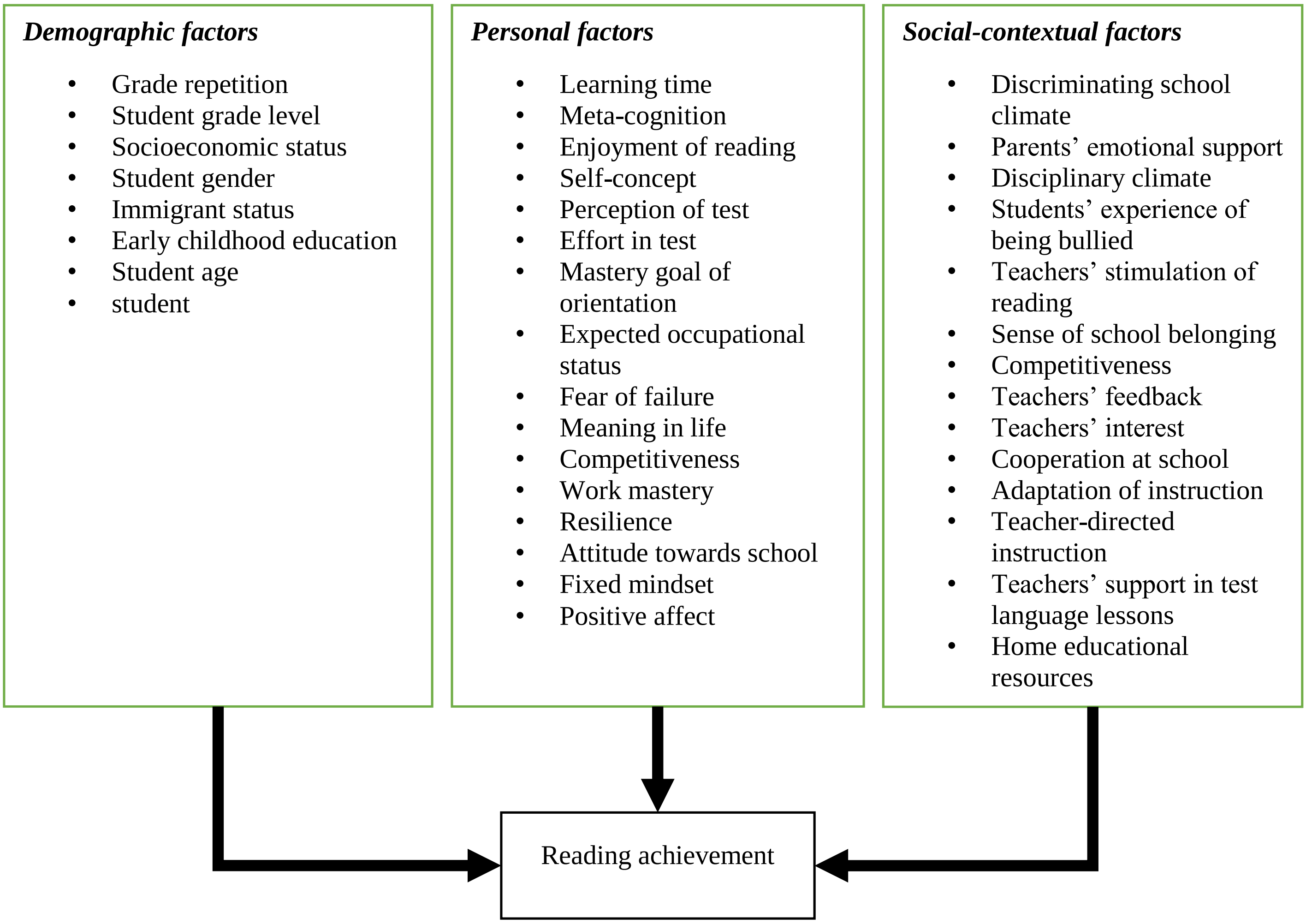

Figure 2

Demographic, personal, and social-contextual variables ranking by variable importance

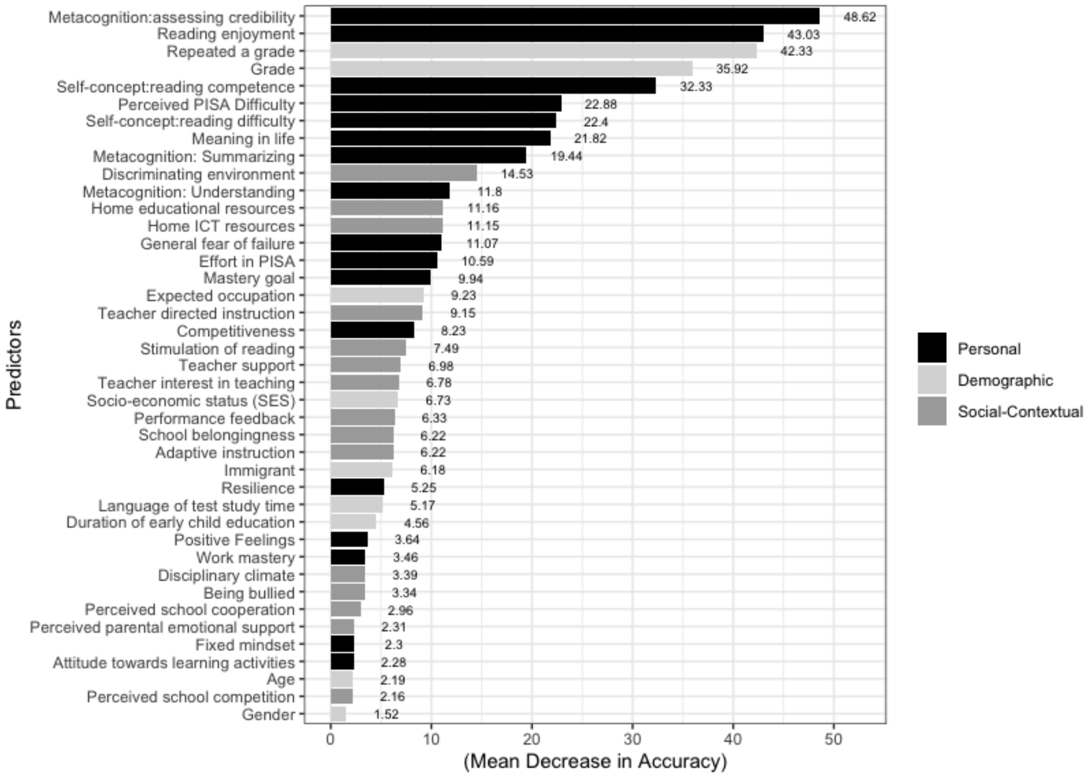

Figure 3

Prediction performance of models with incremental number of predictors

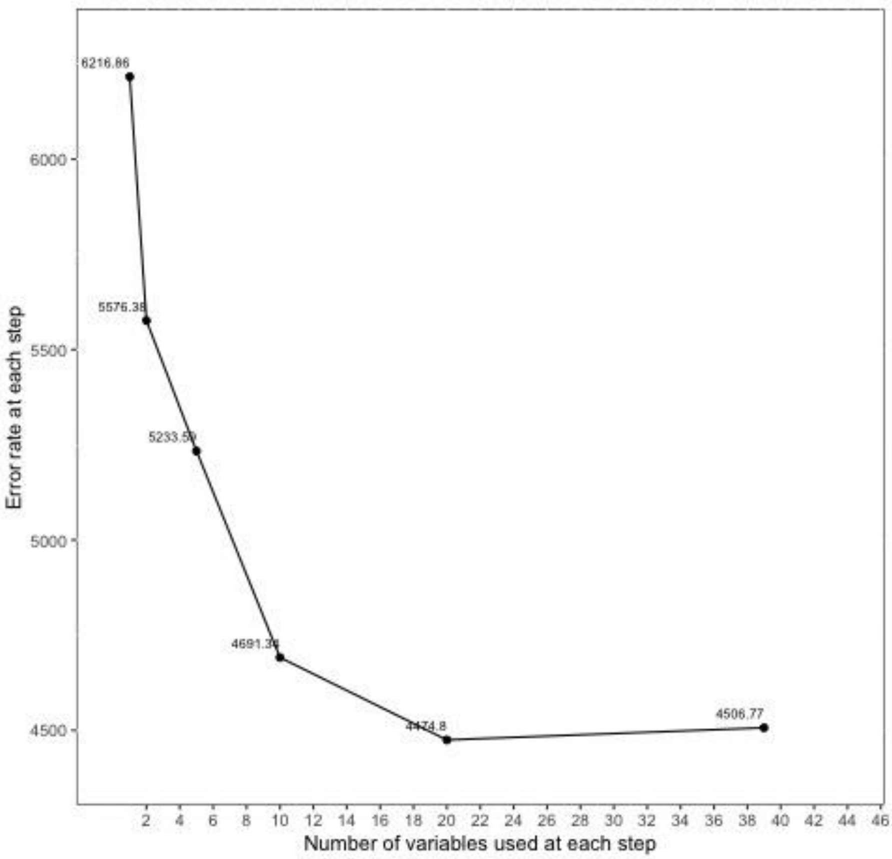

Figure 4

Final variable importance ranking of 20 top predictors of reading achievement

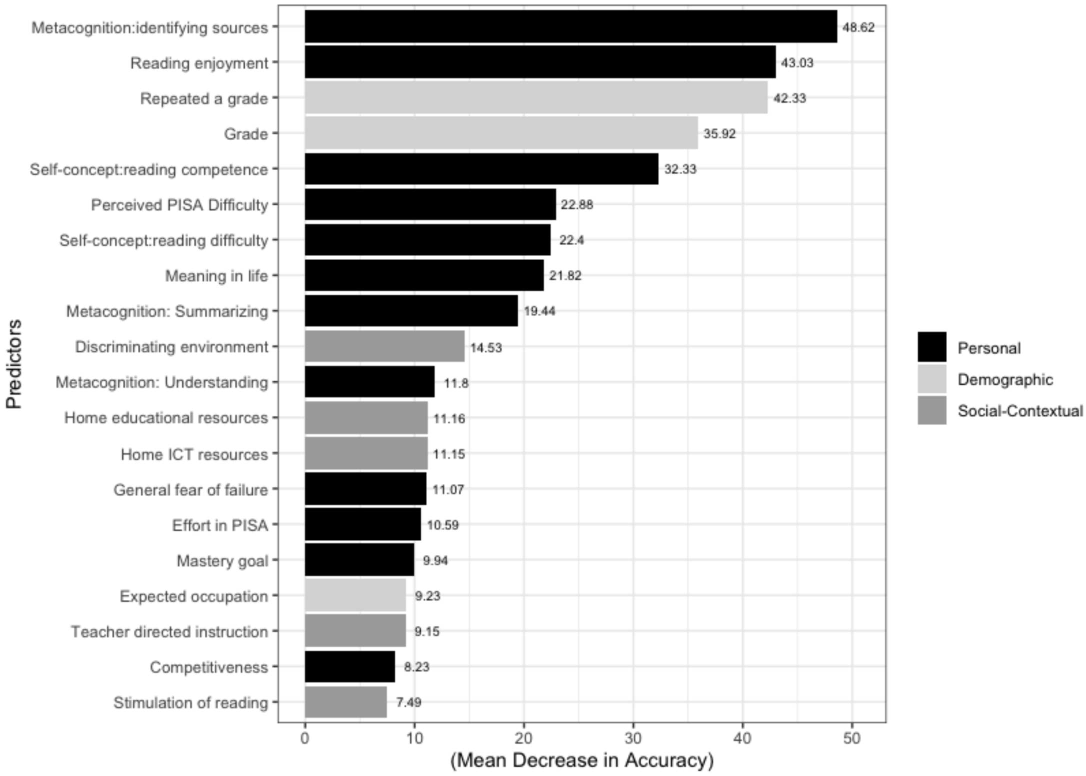

Table 1 .

Bivariate correlations of Top 20 predictors of Macau Students' reading achievement

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1. Reading Achievement																					
2. Metacognition: identifying sources	0.37																				
3. Reading enjoyment	0.33	0.17																			
4. Repeated a grade	-0.42	-0.17	-0.14																		
5. Grade	0.44	0.19	0.11	-0.75																	
6. Self-concept: reading competence	0.31	0.16	0.67	-0.10	0.07																
7. Perceived PISA Difficulty	-0.25	-0.13	-0.28	0.08	-0.07	-0.40															
8. Self- concept: reading difficulty	-0.22	-0.14	-0.28	0.06	~~-0.04~~	-0.35	0.50														
9. Meaning in life	-0.13	-0.07	0.11	0.06	-0.06	0.14	-0.09	-0.07													
10. Metacognition: Summarizing	0.30	0.29	0.18	-0.19	0.20	0.13	-0.10	-0.08	-0.04												
11. Discriminating environment	-0.18	-0.13	-0.12	0.08	-0.06	-0.07	0.08	0.07	0.01	-0.09											
12. Metacognition: Understanding	0.24	0.20	0.17	-0.14	0.17	0.08	-0.05	-0.05	~~0.00~~	0.40	-0.07										
13. Home educational resources	0.22	0.07	0.21	-0.17	0.18	0.22	-0.15	-0.11	0.12	0.10	~~-0.02~~	0.12									
14. Home ICT resources	0.13	0.06	~~0.04~~	-0.11	0.13	0.04	-0.09	-0.09	0.06	0.06	~~0.00~~	0.04	0.38								
15. General fear of failure	0.12	0.10	~~-0.02~~	-0.06	0.08	~~-0.04~~	0.11	0.12	-0.12	~~0.03~~	~~0.02~~	0.06	~~0.01~~	~~0.03~~							
16. Effort in PISA	0.19	0.06	0.15	-0.07	0.08	0.15	-0.08	-0.06	0.09	0.04	-0.08	0.04	0.11	0.04	~~0.03~~						
17. Mastery goal	0.19	0.10	0.31	-0.14	0.11	0.28	-0.14	-0.09	0.22	0.16	-0.11	0.15	0.20	0.08	0.07	0.13					
18. Expected occupation	0.19	0.11	0.12	-0.17	0.15	0.10	-0.07	~~-0.03~~	~~0.02~~	0.09	-0.05	0.08	0.12	0.11	~~0.04~~	0.09	0.18				
19. Teacher directed instruction	~~0.00~~	~~0.02~~	0.13	~~0.00~~	~~-0.02~~	0.14	-0.06	-0.06	0.13	0.04	-0.16	0.04	0.13	~~0.03~~	~~0.01~~	0.06	0.22	~~0.02~~			
20. Competitiveness	0.11	0.06	0.11	-0.06	0.06	0.16	-0.11	-0.07	0.21	0.10	~~0.01~~	0.09	0.16	0.06	0.09	0.12	0.30	0.06	0.12		
21. Stimulation of reading	0.09	0.07	0.19	~~-0.03~~	~~0.00~~	0.18	-0.10	-0.08	0.18	0.09	-0.15	0.09	0.17	0.06	~~0.00~~	0.07	0.27	0.07	0.50	0.18	
Mean	537.78	-0.09	0.30	-	-	-0.41	0.39	0.14	-0.24	-0.27	-0.19	-0.08	-0.15	-0.30	0.47	8.25	-0.22	67.32	0.04	0.14	-0.04
SD	86.85	0.99	0.84	-	-	0.82	0.86	0.90	0.89	0.93	0.82	0.96	1.03	0.80	0.85	1.40	0.91	16.59	0.92	0.80	0.88

Note: Correlation coefficients were estimated using 10 plausible values of reading achievement. *r* coefficients in regular type fonts are statistically significant at $p<.05$; *r* coefficients in strikethrough fonts are non-statistically significant.

Supplementary file I

Table S1.

Variable description, descriptive statistics, and bivariate correlations for overall 41 variables

	Variable Name	Description	*M*	*SD*	*r*
Dependent variable					
	PVREAD1-PVREAD10	Plausible values of student scores in reading	537.78 (1.40)	86.85 (1.08)	-
A.	Demographic Factors				
1	***REPEAT***	Grade Repetition. Values: 0- Did not repeat a grade, 1-Repeated a grade	-	-	-0.42*** (0.02)
2	***GRADE***	Student International Grade.			0.38*** (0.01)
3	SES	Socio-economic status. Index of economic, social, and cultural status (ESCS) derived by PISA from other composite variables: parents' highest educational status, highest occupational status, and general wealth.	-0.66 (0.01)	0.87 (0.01)	0.21*** (0.02)
4	GENDER	Student (Standardized) Gender. Recoded to 0-Male (47%), 1-Female (53%)	-	-	0.11*** (0.02)
5	IMMIG	Index Immigration status. Variable derived from students and student's parents' country of birth. Values: 1 – Native; 2 – First generation; 3 – Third Generation	-	-	0.10*** (0.02)
6	DURECEC	Duration in early childhood education and care. Values: 0-less than a year; 8 – at least 8 years)	-	-	-0.06** (0.02)
7	AGE	Age	15.83 (0.04)	0.29	0.04*
B.	Personal Factors				
8	LMINS	Learning time (minutes per week) for language of test	249.04 (0.93)	54.14 (2.45)	0.00 (0.03)
9	***METASPAM***	Meta-cognition: assess credibility. Question: "In your opinion, how appropriate are the following strategies in reaction to this email?" Sample Item: "Check the sender's email address". Choices: 1-Not appropriate at all to 6- Very appropriate.	-0.09 (0.02)	0.99 (0.01)	0.37*** (0.02)
10	***JOYREAD***	Joy of reading. Sample item: "Reading is one of my favourite hobbies." Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.83	0.30 (0.01)	0.84 (0.01)	0.33*** (0.02)
11	***SCREADCOMP***	Self-concept of reading: Perception of reading competence. Sample Item: "I am able to understand difficult texts." Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.76	-0.41 (0.02)	0.82 (0.01)	0.31*** (0.02)

12	***METASUM***	Meta-cognition: summarising. Students were asked: “How do you rate the usefulness of the following strategies for understanding and memorising the text?” Sample item: “I summarise the text in my own words.” Choices: 1-Not useful at all to 5 – Very useful.	-0.27 (0.02)	0.93 (0.02)	0.30*** (0.02)
13	***PISADIFF***	Perception of difficulty of the PISA test. Sample Item: “Many texts were too difficult for me.” Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.80	0.39 (0.02)	0.86 (0.01)	-0.25*** (0.02)
14	***UNDREM***	Meta-cognition: understanding and remembering.	-0.08 (0.02)	0.96 (0.01)	0.25*** (0.02)
15	***SCREADDIFF***	Self-concept of reading: Perception of difficulty in reading. Sample Item: “I have to read a text several times before completely understanding it.” Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.77	0.14 (0.02)	0.90 (0.02)	-0.22*** (0.02)
16	***EFFORT1***	Effort given in PISA How much effort did you put into this test? Scale: 1 - 10	8.25 (0.03)	1.40 (0.03)	0.19*** (0.02)
17	***MASTGOAL***	Mastery goal orientation. Sample item: “My goal is to understand the content of my classes as thoroughly as possible.” Choices: 1-Not at all true of me to 5 – Strongly true of me. Scale reliability =.84	-0.22 (0.02)	0.91 (0.01)	0.18*** (0.02)
18	***BSMJ***	Student’s expected occupational status. One item question: “what kind of job do you expect to have when you are about 30 years old”. PISA recoded the answer into indices. Higher score means higher expected job.	67.32 (0.26)	16.59 (0.24)	0.17*** (0.02)
19	***GFOFAIL***	General fear of failure. Sample Item: “When I am failing, I worry about what others think of me.” Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.79	0.47 (0.02)	0.85 (0.01)	0.12*** (0.02)
20	***EUDMO***	Eudaemonia: meaning in life. Sample Item: “I have discovered a satisfactory meaning in life.” Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.83	-0.24 (0.02)	0.89 (0.01)	-0.12*** (0.02)
21	***COMPETE***	Competitiveness. Sample Item: “I enjoy working in situations involving competition with others.” Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.71	0.14 (0.02)	0.80 (0.01)	0.11*** (0.02)
22	***WORKMAST***	Work mastery. Sample Item: “If I am not good at something, I would rather keep struggling to master it than move on to something I may be good at.”	0.02 (0.01)	0.84 (0.01)	0.09*** (0.02)

		Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.66			
23	RESILIENCE	Resilience. Sample item: “When I’m in a difficult situation, I can usually find my way out of it.” Choices: 1-Not at all true of me to 5 – Strongly true of me. Scale reliability =.72	-0.36 (0.01)	0.82 (0.01)	0.06** (0.02)
24	ATTLNACT	Attitude towards school: learning activities. Sample Item: “Trying hard at school is important.” 1-Strongly agree to 4 – Strongly disagree. Scale reliability =.81	-0.35 (0.01)	0.9 (0.01)	0.06*** (0.02)
25	FIXED MINDSET	Agree: Your intelligence is something about you that you can't change very much. Choices: 1-Strongly disagree to 4 – Strongly agree.	2.51 (0.01)	0.88 (0.01)	-0.05** (0.02)
26	SWBP	Subjective well-being: Positive affect. Question: “Thinking about yourself and how you normally feel: how often do you feel as described below?” Sample Item: “Happy”. Choices: 1-Never to 4 – Always. Scale reliability =.85	-0.05 (0.02)	1.03 (0.01)	0.03 (0.02)
C.	Social-Contextual Factors				
27	***DISCRIM***	Discriminating school climate. Measures the absence of stereotypes, prejudice, and discrimination. Sample Item: “They have lower academic expectations for students of some cultural groups.” Choices: 1- To none or almost none of them to 4 – to all or almost all of them. Scale reliability =.83	-0.19 (0.01)	0.82 (0.01)	-0.18*** (0.02)
28	EMOSUPS	Parents' emotional support perceived by student. Sample Item: “My parents support my educational efforts and achievements.” Choices: 1-Strongly disagree to 4 – Strongly agree. Scale reliability =.86	-0.36 (0.02)	0.91 (0.01)	0.11*** (0.02)
29	DISCLIMA	Disciplinary climate in test language lessons. Sample Item: “There is noise and disorder.” Choices: 1 – Every lesson to 4 – Never or hardly ever. Scale reliability =.83	0.16 (0.01)	0.79 (0.01)	0.09*** (0.02)
30	BEINGBULLIED	Student's experience of being bullied. Sample Item: “Other students left me out of things on purpose” Choices: 1 – Never or almost never to 4 – Once a week or more. Scale reliability =.72	0.00 (0.02)	0.96 (0.01)	-0.09*** (0.02)
31	***STIMREAD***	Teacher's stimulation of reading engagement perceived by student. Sample Item: “The teacher encourages students to express their opinion about a text.” Choices: 1 – Never or hardly	-0.04 (0.01)	0.88 (0.01)	0.09*** (0.02)

		ever to 4 – In all lessons. Scale reliability =.84			
32	BELONG	Sense of belonging to school. Sample Item: “I feel like an outsider (or left out of things) at school.” Choices: 1- Strongly agree to 4 – Strongly disagree. Scale reliability =.79	-0.39 (0.01)	0.68 (0.01)	0.05* (0.02)
33	PERCOMP	Perception of competitiveness at school. Sample Item: “It seems that students are competing with each other” Choices: 1 – Not at all true to 4 – Extremely true. Scale reliability =.85	0.18 (0.02)	0.95 (0.01)	0.05* (0.02)
34	PERFEED	Perceived feedback. Sample item: “The teacher gives me feedback on my strengths in this subject.” Choices: 1 – Never or almost never to 4 – Every lesson or almost every lesson. Scale reliability =.82	-0.14 (0.01)	0.84 (0.01)	-0.05* (0.02)
35	TEACHINT	Perceived teacher's interest. Sample item: “It was clear to me that the teacher liked teaching us.” Choices: 1- Strongly disagree to 4 – Strongly agree. Scale reliability =.85	-0.10 (0.01)	0.81 (0.01)	0.04 (0.02)
36	PERCOOP	Perception of cooperation at school. Sample item: “It seems that students are cooperating with each other.” Choices: 1 – Not at all true to 4 – Extremely true. Scale reliability =.90	0.09 (0.02)	0.93 (0.01)	0.03 (0.02)
37	ADAPTIVITY	Adaptation of instruction. Sample Item: “The teacher adapts the lesson to my class’s needs and knowledge. Choices: 1 – Never or almost never to 4 – Every lesson or almost every lesson. Scale reliability =.74	-0.23 (0.01)	0.84 (0.01)	0.01 (0.02)
38	***DIRINS***	Teacher-directed instruction. Sample Item:” The teacher asks questions to check whether we have understood what was taught.” Choices: 1 – Every lesson to 4 – Never or hardly ever. Scale reliability =.80	0.04 (0.02)	0.92 (0.01)	0.00 (0.02)
39	TEACHSUP	Teacher support in test language lessons. Sample Item:” The teacher gives extra help when students need it.” Choices: 1 – Every lesson to 4 – Never or hardly ever. Scale reliability =.88	-0.04 (0.01)	0.87 (0.01)	0.00 (0.02)
40	***HEDRES***	Home educational resources	-0.15 (0.02)	1.03 (0.01)	0.22*** (0.02)
41	***ICTRES***	ICT resources available at home.	-0.30 (0.01)	0.80 (0.01)	0.13*** (0.02)

Note: Scale reliability were provided by PISA 2018 technical manual (OECD, 2018); variables that are bolded and italicized refer to the top variables identified by the machine learning algorithm as top predictors of reading achievement.
*** p<.001 **p<.01 *p<.05

Table S2.

Hierarchical linear model of top 20 predictors of Macau reading achievement

Predictors	Estimate	SE	β	Std SE
Grade	18.09***	2.56	0.16	0.02
Metacognition : identifying sources	14.43***	1.38	0.16	0.02
Meaning in life	-12.23***	1.51	-0.13	0.02
Self-concept : reading competence	11.99***	2.19	0.11	0.02
Repeated a grade	-19.72***	3.98	-0.11	0.02
Perceived PISA Difficulty	-8.92***	1.85	-0.09	0.02
Reading enjoyment	8.67***	2.21	0.08	0.02
Discriminating environment	-7.38***	1.69	-0.07	0.02
Metacognition: Understanding	6.79***	1.51	0.07	0.02
Effort in PISA	4.15***	0.97	0.07	0.02
General fear of failure	6.14***	1.51	0.06	0.01
Self-concept : reading difficulty	-4.82**	1.54	-0.05	0.02
Teacher directed instruction	-4.47**	1.61	-0.05	0.02
Metacognition: Understanding	3.51*	1.41	0.04	0.02
Expected occupation	0.21*	0.08	0.04	0.02
Competitiveness	3.00	1.61	0.03	0.01
Home educational resources	2.11	1.46	0.03	0.02
Stimulation of reading	1.74	1.76	0.02	0.02
Home ICT resources	1.47	1.76	0.01	0.02
Mastery goal	1.32	1.53	0.01	0.02
Intercept	320.28***	26.09	-0.10	0.05
Random Effects				
Level 2 Variance	441.03			
Level 1 Variance	3796.29			
Variance explained at Level 2	0.39			
Variance explained at Level 1	0.35			
Total Variance Explained	0.36			
Unconditional Intraclass Correlation Coefficient	0.16			
Conditional Intraclass Correlation Coefficient	0.15			

Note. To account for the nesting of students within schools, the study employed hierarchical linear modelling (HLM) as a supplementary analytic method. Sampling weights and replicates were used to estimate standard errors. Parameter estimates were computed for individual plausible values for reading and then aggregated to yield the final parameter estimate (for the analysis procedures please see OECD, 2009). Parameter estimates are sorted by absolute value of standard estimates (β). This model was able to explain lesser amount of variance in reading achievement compared to the random forest regression model.

***p<.001, **p<.01, *p<.05

Supplementary Material II

Studies that compared the performance of different machine learning algorithms in predicting student outcomes have been well documented (e.g., Costa-Mendes et al., 2021; Fahd et al., 2021; Guo et al., 2016; Martínez-Abad & Chaparro Caso López, 2016). In a recent systematic review of machine learning, some ensemble methods such as decision trees, random forests, and boosting were found the most frequently used in the contemporary literature (Fahd et al., 2021). They are based on classification and regression trees but have been implemented in conjunction with other techniques (e.g., random under-sampling techniques) to improve their performance accuracy, which led them to be considered as "state-of-the-art" machine learning techniques (Sagi & Rokach, 2018).

Furthermore, specific algorithms that implement such techniques also dominate the literature such as Random Forest (random forest algorithm) and Gradient and Extreme Gradient Boosting (boosting algorithms) (Fahd et al., 2021). Meanwhile, various bagging algorithms are also commonly used. Machine learning algorithms are generally considered to have superior prediction performance over classical statistical methods (Yarkoni & Westfall, 2017). A common feature of machine learning that is not usually done in classical statistical methods is the cross-validation procedure that uses a subsample to "train" the model for prediction and uses an unobserved subsample to "test" prediction accuracy.

In this section, we tested four machine learning algorithms that utilized an ensemble method (i.e., Random Forest, Gradient and Extreme Gradient Boosting, Extra Tree, and Treebag) using the regression model as indicated in the main manuscript. Random Forest (RF), Gradient and Extreme Gradient Boosting (XGBoost), Extra Tree (ET), and Treebag are four different state-of-art tree-based ensemble methods. Random Forest algorithm starts at the root node of a tree for all data and estimates each predictor variable to see how it separates two different nodes (Breiman, 2001). XGBoost is a machine learning approach to scale up tree boosting algorithms that uses a method of gradient descent to optimize the loss function (Cui et al., 2017). Extra trees algorithm randomly selects the best feature along with the corresponding value for splitting the node (John et al., 2016). Different from RF using a bootstrap replica, ET uses the whole training dataset to train each regression tree (Geurts et al., 2006). Treebag algorithm combines a multitude of decision trees via bagging (Breiman, 2001). Treebag uses a random selection of features for the best split at each node (Cho et al., 2021).

We compared their prediction performance using RMSE and MAE accuracy metrics as well as the variance explained by the predictive model and chose the algorithm with the smallest error rate and the highest explanatory percentage of variance. We used *caret* package in R (Kuhn, 2022; R Core Team, 2019) to train the prediction model. We used a 10-fold cross-validation procedure (training the model in 10 rounds) for all the algorithms using 75% of the total sample (n=2,979). Only default parameters were used and whenever tuning parameters are required, we set the tuning iteration also at 10 rounds for all the algorithms for a common baseline. After the cross-validation, we compared the unseen 25% of our sample (test set) to test the final prediction accuracy.

Table S4. Comparison of Ensemble Machine Learning Algorithm Prediction Performance

Machine Learning Language	Training set (n=2,234)			Test Set (n=745)
	% Variance Explained (R^2)	Root Mean Square Error	Mean Absolute Error	Root Mean Square Error
Extreme Gradient Boosting (XG BOOST)	37.33	72.56	57.65	68.45
Random Forest	**43.50**	**66.28**	**52.82**	67.73
Extra Tree	42.37	66.76	53.09	67.56
TreeBag	35.01	70.29	55.91	71.22
Gradient Boosting	39.66	69.21	55.04	69.14

Note. The highest R^2, and lowest values of RMSE and MAE are written in bold.

Given the “quality” prediction performance of machine learning algorithms, what remains a challenge for researchers is the selection of appropriate algorithms in their analyses. Researchers usually use accuracy metrics such as Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) as benchmarks for selecting one among competing algorithms. The results were shown in Table S4, suggesting that Random Forest performed better in the training set with the highest percentage of variance explained (R^2 = 43.5%), and predictive accuracy (RMSE = 66.28 and MAE =52.82). Extra Tree comes in second for having the lowest error rate and highest R^2. Extra Tree has a slightly better performance than Random Forest in terms of accuracy in the final prediction performance (using the unseen sample). However, results also showed that given the slight differences, Random Forest model can explain more variance than the model using Extra Tree. Some studies suggested that XG BOOST outperformed Random Forest (Costa-Mendes et al., 2021). However, XG BOOST's performance is dependent on hyper-parameter tuning. Hence, as far as performing within the default parameters (which was used in this section), studies suggested that Random Forest is slightly better in terms of generalizing to out-of-bag sample (Martínez-Muñoz et al., 2019).

Table S5 shows whether the top 20 variables were also detected by the other machine learning algorithms. We found that most top 10 predictors were detected by all four algorithms. Although XG BOOST only detected three of the top 11-20 factors, most of the key factors were identified by Extra Tree and TreeBag. The findings of the four algorithms were broadly similar, which further verified the results of this study.

Table S5. Top 20 variables among four different machine learning algorithms.

Top 20 factors in Random Forest	Factor Ranking			
	Random Forest	XG BOOST	Extra Tree	TreeBag
METASPAM	1	✓	✓	✓
JOYREAD	2	✓	✓	✓
REPEAT	3	✓	✓	✓
GRADE	4	✓	✓	✓
SCREADCOMP	5	NA	✓	✓
PISADIFF	6	✓	✓	✓
SCREADDIFF	7	✓	✓	✓

EUDMO	8	✓	✓	✓
METASUM	9	✓	✓	✓
DISCRIM	10	✓	✓	✓
UNDREM	11	NA	✓	✓
HEDRES	12	✓	✓	✓
ICTRES	13	NA	✓	✓
GFOFAIL	14	✓	✓	NA
EFFORTINPISA	15	✓	✓	✓
MASTGOAL	16	NA	✓	NA
BSMJ	17	NA	✓	NA
DIRINS	18	NA	NA	✓
COMPETE	19	NA	✓	NA
STIMREAD	20	NA	NA	NA

Note. ✓ means that the variable was also identified as among the top 20 most important variables by the alternative machine learning algorithms. NA means that the variable was not identified as a top 20 predictor by the corresponding algorithm.

Reference

Costa-Mendes, R., Oliveira, T., Castelli, M., & Cruz-Jesus, F. (2021). A machine learning approximation of the 2015 Portuguese high school student grades: A hybrid approach. *Education and Information Technologies*, *26*(2), 1527-1547. https://doi.org/10.1007/s10639-020-10316-y

Fahd, K., Venkatraman, S., Miah, S. J., & Ahmed, K. (2021). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies,27*, 1-33. https://doi.org/10.1007/s10639-021-10741-7

Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2016). Predicting students performance in educational data mining. *2015 International Symposium on Educational Technology, ISET 2015*, 125-128. https://doi.org/10.1109/ISET.2015.33

Kuhn, M. (2022). *Package "caret" : Classification and Regression Training* (6.0-91). The Comprehensive R Archive Network.

Martínez-Abad, F., & Chaparro Caso López, A. A. (2016). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, *28*(1), 39-55. https://doi.org/10.1080/09243453.2016.1235591

Martínez-Muñoz, G., Bentéjac, C., & Csörgo, A. (2019). *A Comparative Analysis of XGBoost*. https://www.researchgate.net/publication/337048557

R Core Team. (2019). *R: A language and environment for statistical computing* (3.6.2). R Foundation for Statistical Computing. https://www.r-project.org/

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), 1-18. https://doi.org/10.1002/widm.1249

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in Psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Breiman L. (2001) Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Cho, S. Y., Kim, S. H., Kang, S. H., Lee, K. J., Choi, D., Kang, S., ... & Chae, I. H. (2021). Pre-existing and machine learning-based models for cardiovascular risk prediction. Scientific reports, 11(1), 1-10.

Cui, Y., Cai, M., & Stanley, H. E. (2017). Comparative analysis and classification of cassette exons and constitutive exons. BioMed Research International, 2017, 1-9. https://doi.org/10.1155/2017/7323508

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.

John, V., Liu, Z., Guo, C., Mita, S., & Kidono, K. (2015). Real-time lane estimation using deep features and extra trees regression. In Image and Video Technology (pp. 721-733). Springer.

Cover latter

June 13, 2022
Dr. David MÚÑEZ
Associate Editor
Journal for the Study of Education and Development

Dear Dr. MÚÑEZ,

We are grateful for the opportunity to revise and resubmit manuscript RIYA-2022-0007R1, titled "What explains Macau students' achievement? An integrative perspective using a machine learning approach" to the *Journal for the Study of Education and Development.*

Thank you for your insightful comments that enabled us to improve the quality of our manuscript. We included a point-by-point response to these comments, which is in the attached file "Response to comments". All the comments were properly addressed, and the changes were *highlighted in red color in the main text*. We summarized the main revisions below:

1. **Literature review and the present study**
   In the section on the theoretical framework, we have revised the paragraphs to be more universal and not related to the current study to avoid the interruption of the reading flow. Moreover, we moved and integrated the "Machine learning" section into the methodology part. In addition, we indicated that this study is exploratory in nature in "The Present Study".

2. **Results**
   It was suggested that HLM results could not test the robustness of the RF result by the editor. We agreed with this comment. Thus, the conclusion of HLM results was revised. We addressed HLM was used to complement random forest outcomes.

3. **Discussion**
   We first discussed the influences of personal factors, which were the most relevant findings in this study.

4. **Whole manuscript**
   We have checked and corrected the grammar and typos of the full text.

We believe the comments and feedback have resulted in great improvement to our submission. Please find attached the revision and resubmission.

Warm regards,
Authors

Response to comments

General comments:

"Now I have received a response from one of the experts who revised the original version of the manuscript. The reviewer has acknowledged that the manuscript has been improved and that the revised version addresses many of his/her concerns. Given the delay in processing the original manuscript and the nature of the comments that were raised by the reviewers, I am willing to make an editorial decision to avoid additional delays (unfortunately, recruiting reviewers for this manuscript has been more challenging than expected). Based on those comments, as well as the changes that have been implemented in the revised manuscript, and my own reading, I am happy to support your manuscript for publication in JSED (pending of some issues that must be solved beforehand; see my comments below)."

> **Author's response:** We appreciated your comments on this manuscript. Your comments raised are instrumental for us in improving the quality of our manuscript. Please find below a detailed response of how we attended to address the comments.

Comment 1:

"It is probably a matter of style, but I would rephrase the paragraph that was included in the revised manuscript regarding PSCF (line 56 onwards) if the authors still want that paragraph to be part of the Introduction section. I think the reasons that are provided to adopt/support such framework are universal and do not relate to the current study, exclusively. In other words, there is no need to mentioning that "We used…" or that "In the current study, we used…". As it is, it interrupts the flow of that section (it seems that such paragraph is part of "The current study" section). "

> **Authors' response:** Thank you for this suggestion. We have removed the statements, such as "We used", and revised the paragraph about adopting the PSCF. Please refer to page 3, lines 56-72:
>
> "To best understand these factors, Lee and Shute's (2010) developed a personal and social contextual framework (PSCF) to identify the most important factors that were associated with academic achievement and compare their relative importance against each other. In their study, constructs influencing students' academic performance can be categorized into personal and contextual factors.
>
> The PSCF framework has the following advantages for exploring key factors of students' achievement. First, it highlighted that students' learning was influenced by several distinct factors working in concert with each other (Lee & Shute, 2010). Second, it was established based on a comprehensive review of factors that impinge on academic achievement. The authors systematically searched and reviewed studies with strong empirical findings in terms of students' academic achievement at the K12 level. Third, it was an extensive and expandable framework that integrated many theories such as social cognition, self-regulation, metacognitive processes, engagement, and other constructs of interest to educators. It was open-ended and relatively broad so that it covered a broad range of factors compared to more specific theoretical frameworks which focus on a limited set of factors."

Comment 2:

"I do not particularly agree with the methodological approach. I acknowledge the advantages of MLearning but still think that it is a fishing method that does not help in the behavioural sciences field. Mainly, because it simply contributes to the noise that extant hypothesis testing methods generate. Unfortunately, it does not help to disentangle causality in the phenomena that are evaluated. In any case, I respect the opinion of the authors. Because the current paper is not methodological in nature (as mentioned by the authors in their responses to reviewers' comments), I would incorporate the subsection on Machine learning into the "Analytical approach" subsection within the Methods section. Furthermore, I would trim part of the rationale for using that approach, so readers do not get lost in methodological/technical details."

Authors' response: Thank you for this insightful comment. We have moved "Machine Learning" into the method section and trimmed down the reasons for using this approach. Please refer to pages 9-10, lines 208-236:
"Analytic strategy
The Macau PISA 2018 dataset had a low volume of missing data, with only 11 observations having more than 50% missing values. Markov Chain Monte Carlo (MCMC) imputation was conducted via the *mice* package in R to handle the missing data (van Buuren & Groothuis-Oudshorn, 2011). Given the high accuracy rate of MCMC in imputing missing data, all observations were used in the study. The *bifiesurvey* package was used to generate the descriptive statistics and estimate the independent to dependent variable correlation. The preliminary analysis followed the data analysis procedures of PISA (OECD, 2009c).
Machine learning, which is a type of artificial intelligence that can automatically learn and improve from previous information to accurately predict outcomes, was used in this study. It can detect significant relationships, trends, patterns, exceptions, and anomalies that would otherwise go unnoticed (Sumathi & Sivanandam, 2006). Machine learning approaches have been shown to be applicable to the educational context and could help educators make evidence-based interventions accordingly (Chen et al., 2019; Kiray et al., 2015; Sinclair et al., 2021).

Compared to classical statistical approaches, machine learning methods can recognize the multivariate and complex nature of different predictor variables. First, most classical statistical techniques may encounter overfitting problems, when models are incorrectly fitted to adapt to sample-specific noise (Yarkoni & Westfall, 2017). Machine learning approaches minimize overfitting by using k-fold cross-validation, a subsequent model validation method to determine the optimal number of predictors (e.g., Martínez-Abad, 2019; Martínez-Abad et al., 2020). Moreover, the cross-validation procedure in machine learning streamlines a model by selecting variables that have the greatest contribution in predicting an outcome, thereby increasing the model's performance accuracy. Last, machine learning results are not affected by assumptions (e.g., sample size and collinearity) that strongly influence the p-value (Lu & Ishwaran, 2018). In this regard, machine learning can be used as a complementary technique in analyzing large-scale data with a complex combination of variables. "

Comment 3:
"In the present study section, I would indicate that the study is exploratory in nature and does not aim at testing the PSCF model."

Authors' response: We have indicated the exploratory nature of this study in the present study section. Please refer to page 7, lines 168-173:

"Informed by the PSCF framework, we used key demographic, personal, and social-contextual factors to understand how these different sets of variables were associated with reading achievement in Macau. The study did not test the PSCF model itself. It was exploratory in nature using the PSCF to house the study, select factors from the PISA dataset, and classify them into different domains. The research framework was shown in Figure 1."

Comment 4:

"I would encourage the authors to read the manuscript carefully, so verb tenses make sense. There are many instances in the results section in which different verb tenses are combined (e.g., lines 290 to 293)."

Authors' response: Thank you for pointing this out. We have proofread the whole manuscript to keep the consistency of tenses.

Comment 5:

" "the most two important variables" (line 310) can be replaced with "the most relevant variables…" "

Authors' response: Thank you for this comment. We have replaced these words. Please refer to page 13 line 302.

Comment 6:

"When variables are removed from a model, such variables do not create RMSEA (line 313). It can be simply indicated that "removing such variables contributed to model misfit (RMSEA: XXXX)"."

Authors' response: Thank you for the suggestion. We have revised the sentence. Please refer to page 13, lines 303-304:

"Removing Grade repetition and grade contributed to the model misfit with RMSE values of 42.33 and 35.92, respectively."

Comment 7:

"I understand why an HML was performed, but I do not fully agree that the results of such model serve to support the robustness of the original analytical approach. If that is the conclusion, then, why using MLearning if HLM seems the benchmark?"

Authors' response: Thank you for pointing out this. We have removed such a conclusion and only clarified that the effect size of HLM complemented the results of the random forest algorithm. Please refer to page 13, lines 312-314.

"The effect size of each predictor broadly reflected the importance of the variable found by the random forest algorithm, which complemented our primary results. "

Comment 8:

"In the Discussion section, I would discuss the most relevant findings first (the role of personal variables)."

Authors' response: Thank you for the suggestion. We have moved the subsection of discussion about "Personal factors" before "Demographic factors".

Comment 9:

"Line 402 "the" should be "The"."

Authors' response: Thank you for this kind reminder. We have revised the typo and checked the whole manuscript carefully.