Student well-being and mathematical literacy performance in PISA 2018: a machine-learning approach

Arroyo-Resino D., Constante-Amores, A. Gil-Madrona, P. & Carrillo-López, P.J.

One of the goals of the educational system is to promote the well-being of students due to its associated on their academic performance. This research aims to shed light on the main role of well-being variables (introduced by PISA 2018 for the first time, as far as our knowledge) in the mathematical competence throughout of the PISA 2018 evaluation with a sample of 35,943 Spanish students. The age of the students ranged from 15.333 to 16.333 years, with a mean of 15.836 years (SD = 0.288). Supervised learning techniques such as decision tree methodology, random forest, and a linear hierarchical model have been used throughout this study. The criterion variable was competency performance in mathematics, while the independent variables consisted of a total of 83 items extracted from the student well-being questionnaire. These predictors are grouped into five domains: physical, psychological, material, cognitive and social. We have proved that well-being plays an important role in mathematical understanding in PISA 2018. Specifically, social well-being is the most important variable in our study. To conclude, we observe that social well-being, contextualized in terms of the relationships that the students maintain with their teachers, peers and families, plays a detrimental role in mathematics achievement.

Keywords: PISA; mathematical literacy; performance; well-being; machine learning.

Introduction

The competence profile achieved by students at the end of compulsory education is a concern today, since a low level of competence can lead to a low level of student qualifications in the future, thus negatively affecting the student themselves and society in general (Arroyo-Resino et al., 2019). In aggregate terms, an education system that leaves a large number of its participants behind shows a failure to mobilise their talents and leads to a potential loss of human capital; with the economic (lower long-term growth) and social (higher risk of social exclusion) repercussions that this entails (Organisation for Economic Co-operation and Development (OECD, 2017)). From this perspective, putting resources into education results in the long-term economic growth of the country (Sjøberg & Jenkins 2022).

The concept of a competence profile is a multidimensional and lifelong-learning objective to enable individuals to become more globally competent in order to be able to examine local, global and intercultural issues; understand and appreciate different perspectives and worldviews; and interact successfully and respectfully with others and the environment (Izquierdo, 2018). Out of all the existing programmes for the assessment of academic competencies, one of the most widespread is the *Programme for International Student Assessment (PISA)*. The assessment of this programme focuses on reading, mathematical and scientific literacy from the perspective of understanding concepts, the ability to put them into practice and the specific functions of each skill (OECD, 2019).

Specifically, mathematical competence in Spain is the one that has shown the greatest score difference with respect to the OECD in the latest edition of PISA (see Figure 1) (OECD, 2019).

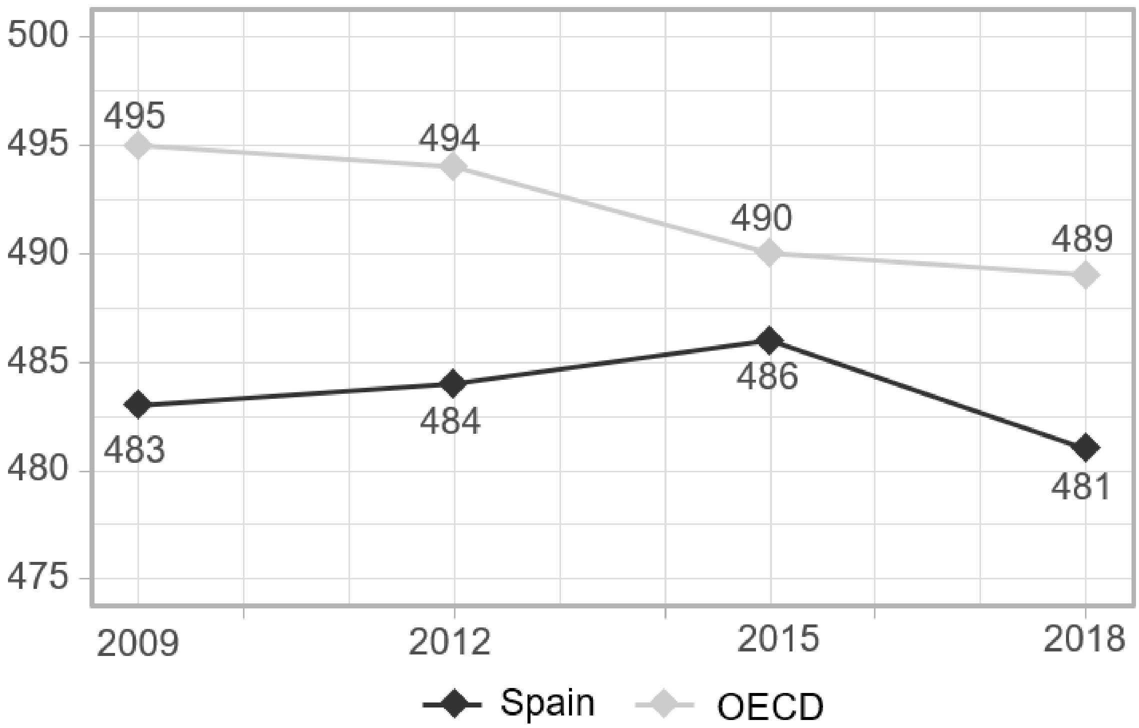

Figure 1. Evolution of estimated mathematical proficiency scores

This competence is the ability of students to formulate, apply and interpret mathematics in different contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena of various kinds (Cordero et al., 2013). The development of this competence at an early age is of utmost importance for its application in everyday life and for the intellectual development of students (Ministry of Education and Vocational Training, 2019; Wang et al., 2023).

In addition, in order to improve the quality and equality of education, *PISA*, since its very first assessment, has been investigating the influence of socio-demographic factors, such as the social, economic and cultural status of families, and other variables such as the use of information and communication technology (ICT), innovation in the teaching and learning process, and family participation, among others, on the level of proficiency in these areas (Sánchez et al., 2019). For example, from the data analysed in PISA 2015, the importance of subjective and emotional dimensions in scientific literacy is clear (Gil et al., 2019). In the most recent edition (2018), *PISA* has included questions on students' social and emotional status, including how satisfied they are with their life, their feelings and their fear of failure, in order to gain a more holistic understanding of students' educational performance and well-being. This optional questionnaire, linked to well-being, sought to connect school life to the broader ecosystem in which students, family, their peers and the community live, with the aim of obtaining information on the holistic development of 15-year-old schoolchildren (OECD, 2019).

In this line, the scientific literature highlights that one of the most analysed personally identifiable variables since the last decade of the 21th century is the well-being of the child and youth population, due to its importance for sufficient mental health, academic achievement and professional success (Gutiérrez et al., 2021; Govorova et al., 2020). Hence, the well-being of the student population is considered a key component of academic performance and, thus, of the socio-economic well-being of the population to which they belong, as several studies show a positive relationship between the two phenomena, such that a high economic status in a given population is characterised by students with a high academic level (Clarke, 2020; Cobo et al., 2017).

PISA 2018 states that student well-being is a dynamic state of psychological, cognitive, material, social and physical functioning and capabilities needed to live a happy and fulfilling life, which persists into adulthood (Spanish Ministry of Education and Vocational Training, 2019). Within this construct, building on the study by Borgonovi & Pál (2016), the OECD (2019) identifies five domains: I) *cognitive well-being*, referring to the knowledge, skills and foundations that students have to participate effectively in today's society, as lifelong learners, effective workers and engaged citizens, II) *psychological well-being*, which includes students' views and assessments of their lives, their commitment to school, and the goals and ambitions they have for their future, III) *physical well-being*, concerning students' health, commitment to physical exercise and adoption of healthy eating habits, IV) *social well-being*, referring to the quality of their social life, including relationships with their family, peers and teachers, and how they perceive their social life at school; and finally, V) *material well-being*, understood as the material resources that enable families to meet their children's needs and schools to support students' learning and healthy development.

Recent research has documented significant and steady declines in adolescent well-being (Boonk et al., 2018). As a result, curriculum policies have introduced changes that promote the well-being of schoolchildren (Clarke, 2020). As an essential part of students' lives, the academic context can promote and facilitate their well-being by including it in the planning and teaching-learning process, thereby connecting many domains and helping to build relationships of trust and respect. However, it could also hinder it depending on the type and number of actions carried out to enrich the social and pedagogical interactions offered to all students (López et al., 2021). Furthermore, the social and economic circumstances that shape students' individual experiences and their family experiences may have an impact on their well-being (Murillo & Hernández, 2020). In several longitudinal studies, the results reveal positive and significant relationships between academic performance and different well-being indicators (physical, cognitive and emotional) (Kleinkorres et al., 2020; Yang et al., 2020). Moreover, in a review of the main psychological and educational theories, using data from PISA 2012, Clarke (2020) shows that, far from being incompatible, adolescents' well-being and achievement can be positively associated. However, he points out that this relationship is not straightforward and requires careful disentangling of the hedonic and eudaimonic components of well-being. In this line of argument, several meta-analyses have shown the existence of a large body of literature on student well-being and academic performance, with empirical research yielding ambiguous results (Bücker et al., 2018; Kaya & Erdem 2021). In turn, the study by Govorova et al. (2020), from a transcultural perspective in 35 OECD countries and using data from PISA 2015, contributes to the general discussion, currently underway, on the definition of well-being and the connection between well-being and academic achievement. Likewise, the systematic review developed by Wang et al. (2023), encourages the development of more specific research to understand the well-being factors that affect academic performance in mathematics in the PISA reports.

In light of the above, the objective of the present research is to study the influence of the variables of the student well-being on mathematical literacy in PISA 2018. This objective is specified as follows:

1. Identify the variables associated with well-being that show the greatest differentiation between individuals who achieved higher and lower scores in the PISA 2018 assessment. This will help in characterizing the profiles of young people with varying levels of mathematical literacy.
2. Identify the student well-being variables that have the greatest influence on mathematical literacy.
3. Once the influence of the socio-economic context has been monitored, determine the student well-being variables that most associated mathematical literacy taking into account the hierarchical structure of the data (level 1: student and level 2: school).

Method

Design

In order to address the study objectives, a secondary analysis of the PISA 2018 database provided by the OECD (2019) was conducted. The methodology of the study is characterised as quantitative with a non-experimental design based on cross-sectional data.

Sample

The sample consists of the Spanish student body (35,943) that participated in the PISA 2018 assessment, where 50% were male and 50% female. The age of the students ranged from 15.333 to 16.333 years, with a mean of 15.836 years (SD = 0.288). Approximately 64% of students were enrolled in state schools, 29% in state-funded schools and 7% in private schools.

Variables

The response variable of the present study is mathematical literacy, which assesses an individual's "capacity to formulate, employ, and interpret mathematics in different contexts" (OECD, 2019, 5). It should be noted that this is made up of ten plausible values as of the PISA 2015 edition (OECD, 2019), which have been created using the response imputation method based on Item Response Theory (OECD, 2018).

The independent variables were drawn from the PISA 2018 student well-being questionnaire (https://acortar.link/WHW74T). Specifically, student responses to the 83 items simple of this instrument were used, which are grouped into five domains: physical, psychological, material, cognitive and social (OECD, 2017).

Procedure and data analysis

Prior to pursuing the different objectives based on the machine learning algorithm, the data were pre-processed, as recommended by Lantz (2013). Firstly, missing values were processed using the multiple imputation by chained equations technique (Rivero, 2011). Secondly, the data were divided randomly into two sets. The first set corresponds to the "training set" used during the learning process (70% of the data). The second set of data is used to validate the model throughout the root-mean-square error (RMSE), and subsequently, a value close to zero indicates a low prediction error (Raschka & Mirjalili 2019). Thirdly, the continuous predictors were centred and scaled to prevent variables with a higher magnitude being of greater importance in the model. Fourthly, independent variables were checked for zero or near-zero variance in order to eliminate those that do not provide information. Nevertheless, no predictors were removed as all of them showed variability.

Once this processing had been carried out, in order to achieve the first study objective of describing the profile of students with the highest and lowest scores in mathematical literacy, the regression tree method was used. This machine learning algorithm was repeated for each of the ten plausible values in order to select the value with the lowest root-mean-square error (RMSE), as in the research of López-Martín et al. (2018). In order to obtain less overfitting, the hyperparameters of the decision trees (maximum depth, minimum number of observations at a terminal node and at each node) were optimised as a whole by means of 10-fold cross-validation (Lantz, 2013). This procedure showed that plausible value number eight has the lowest RMSE, that is, the lowest estimation error (minimum number of observations at each node = 25; maximum depth = 3; minimum number of observations at a terminal node = 100; RMSE = 82.985). With regard to the second objective based on identifying the student well-being variables that have the greatest impact on mathematical literacy, the Random Forest method (Breiman, 2001) was used for each of the plausible values in order to select the plausible value with the lowest RMSE. This algorithm is characterized by being highly efficient to identify the most important variables (Raschka & Mirjalili, 2019). It should be noted that, in the same way as in the previous objective, the hyperparameters here, in this case of random forests, are also optimised as a whole. Out of the various plausible values, number 8 was again selected as it has the lowest RMSE (number of variables = 10; number of trees = 401; maximum depth = 25; RMSE = 75.579). It is observed that the Random Forest method is more accurate than regression trees, as it shows a lower estimation error (RMSE). Once this procedure had been carried out, the first fifteen most important variables were ordered using the permutation statistic. On this basis, there is no statistical criteria to select the predictors that have a significant role in the response variable (Raschka & Mirjalili, 2019).

The third objective is to determine the variables related to student well-being that are linked to their mathematical competence, once the influence of the socio-economic context has been minimised. For this purpose, the nested structure of the educational data (Lee, 2000) was taken into account by estimating different hierarchical linear models (Brown, 1994; Raudenbush & Willms, 1995), considering level 1: student and level 2: school, where the random variance between levels is statistically significant and the Intraclass Correlation Coefficient is higher than 10%, as recommended by Gaviria & Castro (2005). Prior to running the models, the assumption of multicollinearity was checked in two stages. Firstly, we correlated the variables using Spearman's correlation coefficient, as none of the variables met the assumption of normality. The work by Raschka & Mirjalili (2019) states that the relationship is considered to be strong when values are larger than 0.8. Secondly, we checked that the Variance Inflation Factor was lower than 10 and as this was true in all cases1, we concluded that the assumption of multicollinearity had been met.

Three models were estimated. Model 0 or null model, with no predictors. Model 1, where the economic, social and cultural status of both the student and the school are entered, as they are considered control variables in the final model, as recommended by the OECD (2009). And finally, model 2 which includes the fifteen most important variables of student well-being (result of specific objective 2), following the order established by the machine learning algorithm, as has been done in similar studies (Arroyo-Resino et al., 2019; Constante-Amores et al., 2021). In addition, as mentioned above, the influence of the economic, social and cultural status of both the student and the school is monitored due to the impact it may have on student performance (Cordero et al., 2013).

For a proper multilevel analysis, when working with plausible values, the indications in the research conducted by Laukaityte & Wiberg (2017) were followed. Furthermore, in order to analyse the multilevel effect size of each predictor, Cohen'*s d* is calculated in the final model, where a value of around 0.20 indicates a small effect size, values around 0.50 a medium effect size and values around 0.80 and higher a large effect size (Cohen, 1992). Finally, the AIC, BIC (Akaike Information Criterion and Bayesian Information Criterion, respectively) and *Deviance* statistics are used to analyse the misfit of the estimated models, where low values indicate a better fit. The significance of the reduction in misfit is also calculated by subtracting the *deviances* of the compared models and estimating the chi-square distribution of said difference, using the difference between the estimated parameters in each model as degrees of freedom (Cameron & Windmejier 1997).

All analyses were performed using R software, version 4.0.5. Specifically, the following packages were used: *mice* (multivariate imputation by chained equations), *randomForest* (Random Forest), *caret* (machine learning) and *nlm4* (hierarchical linear model).

Results

Specific objective 1: profile of high and low achievers in mathematics

Once the hyperparameters of the regression tree algorithm were optimised as a whole, the student profile was then obtained (see Figure 2).

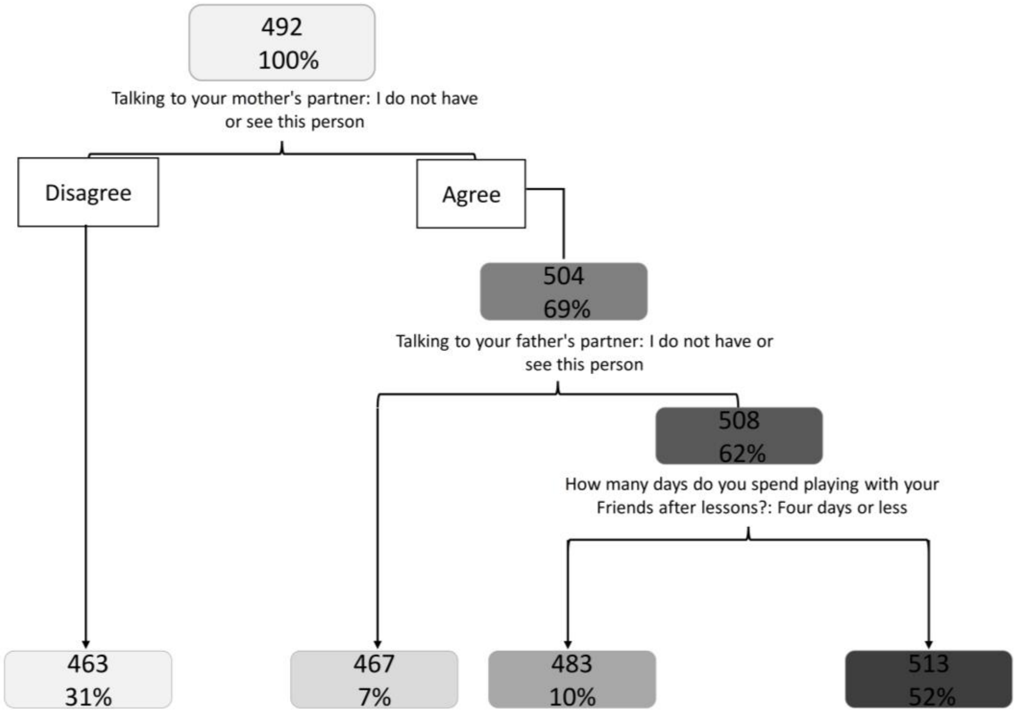

Figure 2. Profile of high and low achievers in mathematics

Students with the highest mathematical literacy score (513 points) are those who do not talk to their mother's or father's partner when something is bothering them (either due to absence or lack of relationship) and, in addition, spend time with their friends four days or less right after school. It should be noted that this profile refers to 52% of the training sample. Nevertheless, the students with the lowest performance in the response variable (463 points) are characterised by the fact that they do not talk to their mother's partner when something is bothering them.

Specific Objective 2: importance of well-being variables that influence mathematical performance

After optimising the random forest hyperparameters as a whole and with the lowest RMSE obtaining plausible value 8, the 15 student well-being variables that most impact mathematical literacy were then obtained (see Table 1).

Table 1. Well-being variables that most impact mathematical literacy

Independent variables	**Importance (permutation)**
How easy is it for you to talk to your mother's partner about things that really bother you?	305.036
How easy is it for you to talk to your father's partner about things that really bother you?	270.396
How many days a week do you usually spend time with your friends right after school?	219.360
Outside of school, during the past 7 days, on how many days did you engage in moderate physical activities?	146.819
How satisfied are you with all the things you have?	131.185
How easy is it for you to talk to other members of your family about things that really bother you?	123.626
How easy is it for you to talk to your teachers about things that really bother you?	119.465
How often do your parents or guardians try to control everything you do?	115.878
Did you feel bored the last time you attended a mathematics class at school?	107.256
How many close friends do you have?	104.007
Did you feel happy the last time you spent time outside your home with your friends?	98.503
Are you satisfied with your health?	95.071
Are you satisfied with the way you use your time?	71.546

I like my look just the way it is	70.128
How many days a week do you attend physical education classes?	64.993

Specifically, the three predictors with the greatest impact are the following: *How easy is it for you to talk to your mother's partner about things that really bother you? How easy is it for you to talk to your father's partner about things that really bother you? And how many days a week do you usually spend time with your friends right after school?* These predictors are related to the student's social well-being dimension.

Specific objective 3: predictors of mathematical literacy

To facilitate the interpretation of the results, we perform a dichotomization of the independent variables of a categorical nature (following the recommendation by Pardo & Ruiz (2013) (see Table 2); then we use the hierarchical-linear models. In the case of the discrete quantitative variables (moderate physical activity, number of Physical Education classes, and number of days spent with your friends after class) they have been dichotomized using the decision tree technique using the CART growth method (Raschka & Mirjalili, 2019).

Table 2. Predictor variables

Variables	**Recoded values**
Economic, Social and Cultural Status (ESCS) of the student	Quantitative variable
Average ESCS of the school	Quantitative variable
Talking to your mother's partner	0 = I don't have or I don't see this person. 1 = I get along with this pearson

Talking to your father's partner	0 = I don't have or I don't see this person. 1 = I get along with this pearson
Number of days spent with friends right after school	0 = Four days or less 1 = Five
Moderate physical activity	0 = Four days or less 1 = Five, six and/or seven days
All the things you have	0 = Not very satisfied or not at all satisfied 1 = Satisfied and totally satisfied
Talking to other family members	0 = I don't have or see this person 1 = Very easy, easy, difficult and very difficult
Talking to the teacher	0 = I don't have or see this person 1 = Very easy, easy, difficult and very difficult
Parental or guardian control	0 = Almost never 1 = Sometimes or almost always
Bored in the last mathematics class	0 = A little or not at all 1 = Quite or extremely
Number of close friends	Quantitative variable
I feel happy when I spend time with my friends	0 = A little or not at all 1 = Quite or extremely

Health	0 = Not very satisfied or not at all satisfied 1 = Satisfied and totally satisfied
Use of time	0 = Not very satisfied or not at all satisfied 1 = Satisfied and totally satisfied
The way I look	0 = Disagree, strongly disagree or I don't have an opinion. 1 = Agree or strongly agree
Number of physical education classes	0 = One to two days 1 = Zero, three or more days

Table 3 shows the results of the hierarchical linear models. Model 0 is the model with no predictors where the largest misfit occurs, model 1 includes the socio-economic context variables, which significantly predict student performance. Specifically, for every point increase in the school's ESCS and average ESCS, there is an increase of 23.101 and 25.122 points in mathematical literacy, respectively. These predictors explain 61% of the variability in mathematical literacy. Model 2 includes the 15 most important variables of student well-being identified in the previous objective. It is noteworthy that all were statistically significant, except *How easy is it for you to talk to other members of your family about things that really bother you?* and *I like my look just the way it is.* This model increases the explanation of the response variable by 6%, where the total percentage of variance explained is 67%. With regard to the coefficients of the predictors of the social well-being dimension that were found to be significant, it can be seen that *students who talk to their mother's and/or father's partner when something is bothering them* score 16.102 and 14.687 points lower in mathematical literacy compared to students who do not have or do not see that person, respectively. However, those who *talk to their teachers when something is bothering them* score 20.757 points higher compared to students who cannot rely on this support. Students who *spend time with their friends five days a week right after school* score 20.494 points lower in mathematics performance compared to those who spend four days or less. A decrease of 0.906 points in the response variable is predicted when the student has an additional close friend. Students who *felt quite or extremely happy when they were with their friends outside their home* scored 21.782 points higher than those who did not feel happy or felt a little unhappy. *Students who are sometimes or almost always controlled by their parents or guardians* score 16.665 points lower on average in the response variable compared to students who are almost never controlled. With regard to physical well-being, *students who engage in moderate activities 5 to 7 days a week* score 15.576 points higher compared to those who engage in moderate physical activities four days or less. Students who *attend physical education class zero, three or more days* score 31.573 points less compared to those who attend between one and two days. Students who are *satisfied or totally satisfied with their health* score 12.006 points higher in mathematical literacy than those who are not satisfied or not at all satisfied.

Regarding material well-being, *students who are satisfied or totally satisfied with all the things they have* score 23.112 points higher on the dependent variable compared to those who are not very satisfied or not at all satisfied. Finally, in relation to psychological well-being, *students who are satisfied or totally satisfied with their use of time* score 12.603 points lower than those who are not very satisfied or not at all satisfied. Those students who mentioned that “they were quite or very bored in the last lesson of math” have 8,548 points less than those who said that were little or not at all. Additionally, students who have a good body image scored 2,356 more points in mathematical competence.

In relation to the multilevel effect size of the final model, it can be seen that the variables with the largest magnitude are *ESCS* (*d* = 0.400), *average ESCS* (*d* = 0.270) and *Number of physical education classes* (*d* = 0.220), where these effects are small (Cohen, 1992). Regarding the overall fit of the models, we observe a reduction in the *deviance* of model 2 (202344.578) with respect to that of the null model (209985.143), resulting in a decrease in the significant unexplained variance ($X^2 = 2668.188$; gl = 17; p = .000) and, thus, we conclude that the student well-being variables selected help to improve the fit of the baseline model. These improvements are also observed in the AIC and BIC information criteria, where, as shown in Table 3, these values decrease as predictors are included in the model.

Finally, regarding the calculation of the intraclass correlation coefficient (ICC), in the null model the proportion of variance due to the school is 15.014%. With the introduction of the student well-being questionnaire variables and the monitoring of the economic, social and cultural status of both the student and the school, the variance is reduced to 5.656%. Therefore, the variables entered in the model explain 9.358% of the variation in mathematical literacy results between schools.

Table 3. Multilevel random-intercept regression models to explain mathematical literacy

	Model 0 (null model)	**Model 1 (+ socio-economic context)**	**Model 2 (+ student well-being)**	**Cohen's *d***
Intercept	489.181 (1.126) ***	493.614 (0.751) ***	463.780 (3.736) ***	

ESCS	23.101 (0.476) ***	18.849 (0.513) ***	0.400
Average ESCS	25.122 (1.497) ***	21.789 (1.320) ***	0.274
Talking to your mother's partner		-16.102 (1.609) ***	0.172
Talking to your father's partner		-14.687 (1.095) ***	0.152
Number of days spent with friends right after school		-20.494 (1.039) ***	0.154
Moderate physical activity		15.576 (0.879) ***	0.200
All the things you have		23.112 (1.692) ***	0.200
Talking to other family members		-1.156 (4.350)	0.022
Talking to the teacher		20.757 (2.933) ***	0.076

Parental or guardian control	-16.665 (0.923) ***	0.190
I was bored in the last mathematics class	-8.548 (0.825) ***	0.116
Number of close friends	-0.906 (0.092) ***	0.178
I feel happy when I spend time with my friends	21.782 (1.516) ***	0.148
Health	12.006 (1.258) ***	0.124
Use of time	-12.603 (0.893) ***	0.166
The way I look	2.356 (0.907)	0.028
Number of physical education classes	-31.573 (1.474) ***	0.220

Intra-school variance	1162.255	453.225	384.563
Inter-school variance	6567.788	6137.290	4437.026
ICC	15.014%	12.675 %	5.656 %
AIC	419858.556	409264.300	368311.980
BIC	420001.750	409306.630	404903.030
Deviance	209985.143	204589.486	202344.578
Percentage of variance explained		61.004%	66.912 %

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Discussion and conclusions

This research intended to study the influence of the student well-being questionnaire variables on mathematical literacy in PISA 2018, as said questionnaire was included for the first time in this edition. This paper is, therefore, considered novel, as it is one of the first studies to comprehensively address the student well-being variables associated with performance.

Specific objective 1: profile of high and low achievers in mathematics

With regard to the first objective, aimed at describing the profile of students with the highest and lowest scores in mathematics, it was found that the profile of students with the highest score in mathematical literacy consists of students who do not have or do not see to their mother's partner or their father's partner (which may be due to the absence of said person) when something is bothering them, and who spend time with their friends four days or less right after school.

It is prescriptive to note that all of these predictors are related to social well-being. One possible explanation for these results may be through achievement goal theory (the ability to work hard to demonstrate competence in educational, sporting and/or family settings) and social comparison theory. In this regard, a longitudinal study by Zhou et al., (2020) with adolescents, examined how achievement goals relate to the subjective well-being of adolescents in the school context. To this end, they addressed a theoretical model that specified that academic social comparisons and self-esteem would serve as mediators in the relationship between achievement goals and well-being. The findings revealed that the domain objectives (For example: "I want to learn as much as possible from this class") showed successive indirect associated on subjective well-being at school through social comparisons, academic comparisons and self-esteem. Therefore, in the absence of close social relationships or negative communication with the mother/father figure, the ability to understand and interact in the social world may be affected and, thus, so may their academic performance (Shirley et al., 2020).

Specific Objective 2: importance of well-being variables that influence mathematical performance

In this sense, it is coherent that, with regard to the second objective, which consisted of identifying the variables that most affect performance in mathematics, the three predictors with the greatest impact were the following: *How easy is it for you to talk to your mother's partner about things that really bother you? How easy is it for you to talk to your father's partner about things that really bother you? And, how many days a week do you usually spend time with your friends right after school?* Also, it is prescriptive to note that all of these predictors are related to social well-being.

The study by Molina-Muñoz et al. (2023) indicates that students' emotions are closely related to their performance in mathematics. That is, the emotional state, corresponding to the individual component of students' psychoemotional well-being, the academic or school component and the social component are related to the score or grade obtained in mathematics.

In this context, adolescence is a critical period of development both psychologically and psychopathologically. During this stage, problems with parents and friends have been shown to increase adolescents' depressed mood (Fiorilli et al., 2019). Specifically, regarding the predictors of depression, self-esteem is the most important, followed by maternal and paternal emotional availability (Babore et al., 2016). A longitudinal study confirmed a reciprocal relationship in parent-child cohesion, self-esteem and academic achievement (Wang et al., 2021), with significant bidirectional relationships observed between self-esteem and well-being at school (Yang et al., 2020). These results can be explained from a socio-ecological perspective. That is, social well-being can be interpreted as the result of an individual's interaction with different social settings; thus, the personal characteristics of a schoolchild, when interacting with their family and educational context, will affect the quality of their well-being. In this case, the family and the school as microsystems offer opportunities to develop the sense of belonging to different meaningful groups (Widlund et al., 2018). In particular, the prediction of adolescents' subjective well-being is heavily related by the direct effect of perceived family support (Gutiérrez et al., 2021).

In this sense, school-based parental involvement acts as a powerful predictor of achievement and school learning environment (Park et al., 2017) as, according to the meta-analysis by Smith et al. (2020), they work together to promote child development through activities that link both environments and contribute to the academic and social and emotional competencies of schoolchildren. An even greater difference was found when the mother provides this support. When parents attend meetings with management and teachers, and participate in extracurricular activities, it influences student performance in different school subjects (Murillo & Hernández, 2020). In this sense, it is important to encourage parents' participation and involvement in the school, such as becoming teaching assistants, outdoor volunteers or school board members (Fleer and Rillero, 1999).

The meta-analysis conducted by Sheridan et al. (2019) examined the effects of family and school interventions on schoolchildren's social-behavioural competence and mental health. It showed that interpersonal and relational processes (that is, communication, collaboration and parent-teacher relationship) and tangible structural elements (that is, involvement at home, behavioural support) bring benefits to these competencies and the mental health of schoolchildren. This interpersonal context may play an important role in promoting general well-being and reducing distress by meeting the key psychological needs for autonomy, competence and relatedness, which would explain the relationship between the number of days per week spent with friends and academic performance (Buzzai et al., 2020; Cosentino & Karwowski, 2019).

Specific objective 3: predictors of mathematical literacy

In relation to the third and final objective, intended to determine the predictor variables of mathematical literacy, it should be noted that students who talk to their mother's and/or father's partner (broken homes) score 16.102 and 14.687 points lower, respectively. However, those who talk to teachers when something is bothering them scored 20.757 points higher. In addition, students who are satisfied with their health, with the use of their time or with all the material things they have score 12-15 points higher. Similarly, those who feel quite or extremely happy when they are with their friends away from home scored 21.782

points higher. Nevertheless, those who feel that they are sometimes or almost always controlled by their parents or guardians score 8.548 points less on average.

These aspects can be explained due to the control of emotions (emotional intelligence and emotional state) as they can affect adolescents' resilience and motivation, in addition to academic performance and the adoption of healthy lifestyle habits (Trigueros et al., 2019; Urrila et al., 2017). In this line of argument, it was shown that adolescents with little parental support are less motivated, with poorer academic performance and higher levels of alcohol consumption (Moral-García et al., 2020). In this sense, several systematic reviews (Cobo et al., 2017; Iglesias-Díaz & Romero-Pérez 2021; Tomé et al., 2021) conclude that well-being can be learned in the classroom as schools play a critical role in positive socialisation, in establishing and maintaining cultural values due to the fact that students spend a large part of their time there. Furthermore, according to the meta-analysis conducted by Tan et al. (2020), the socio-economic level of families has some bearing on these aspects. These findings call for integrated educational policies and practices to foster students' sense of belonging, feeling of integration and contribution to their schools and family relationships (López et al., 2021). Therefore, teachers and parents should cultivate students' well-being by focusing not only on present life satisfaction and positive affect, but also on positive feelings regarding the future, given that they are associated with subsequent academic achievement (Wu et al., 2020).

Moreover, within the domain of physical activity, it is noteworthy that students who engage in moderate physical activities 5 to 7 days a week score 15.576 points higher compared to those who engage in moderate physical activities four days or less. Similarly, students who attend physical education class zero, three or more days score 31.573 points less compared to those who attend between one and two days.

The results obtained are similar to those found in the research conducted by Gómez-Fernández & Albert (2020), and show that physical activity can play an important role in academic performance. The results suggest a positive association between the number of days per week engaging in moderate physical activities and performance in scientific, reading and mathematical literacy. However, more days per week of vigorous physical activities was associated with lower scores in reading and science. Therefore, these authors emphasise that the associated of physical activity on children and young people need to be further investigated. With regard to physical education classes, these results can be explained by the increased motor engagement time in these classes in two sessions per week (55 minutes) compared to schools that choose to teach three sessions per week (45 minutes). In other words, in 45-minute sessions there is less optimisation of motor engagement time, which may prevent increased physiological engagement time, understood as the time in the physical education class in which students work at sufficiently significant intensities so as to produce organic and academic improvements (Rodríguez et al., 2020).

In view of the above and based on the study conducted by Jorge & González (2017), it is necessary to develop research the implementation of education and intervention programmes for family life in order to create changes regarding the way children are raised, disciplinary practices and the attention that parents pay to their children, so as to promote an optimal educational climate for learning, healthy lifestyle habits and the development of skills. It is also clear that there is a need to implement programmes that work on students' emotions through emotional education, which will have positive influence on both students' psychosocial environment and their academic performance (Gil et al., 2019; Govorova et al., 2020).

In line with the study objective, it was observed that the student well-being variables that most affect mathematical literacy are related to social well-being, both in the absence of communication with the mother's or father's partner and in interaction with friends. At the same time, the role of physical education and moderate physical activity as performance predictors should be highlighted. These results are of particular interest to the educational society; family, school and students, as an increase in student well-being may lead to improved mathematical literacy.

Based on these findings, it is necessary to ask what the school can do to promote student well-being. In this regard, the school must primarily improve school climate: approaches to learning and assessment, pedagogy, accountability measures, interpersonal relationships, commitment to social-emotional learning and school connectedness, i.e. healthy relationships with different educational agents such as parents and peer friendships (Clarke, 2020; Gutiérrez et al., 2021). Likewise, intervention programs should be developed within the educational context to improve the emotional intelligence of the parents' partners in order to obtain greater emotional stability in the schoolchildren (Molina-Muñoz et al., 2023; Jungert et al., 2020).

At the same time, the aim should be to increase students' personal growth. And personal growth requires fostering commitment, interest, meaning and purpose. And along with these characteristics would also be self-esteem, optimism, resilience, physical and emotional vitality and self-determination (Gil-Madrona et al., 2019; Gómez-Fernández & Albert, 2020; López et al., 2021).

Faced with this situation, teachers have the enormous challenge and, at the same time, the enormous opportunity to develop well-being in the classroom in favour of students, since no student can be left out of the group's learning for any reason (Govorova et al., 2020; Iglesias-Díaz & Romero-Pérez, 2021).

Finally, it should be noted that, from a methodological point of view, the Random Forest method is more accurate than regression trees, as it shows a lower estimation error (RMSE). This is in line with the studies carried out by Lantz (2013) and Raschka & Mirjalili

(2019). Therefore, the use of the Random Forest algorithm is recommended when working with predictive models.

Limitations and future research

Nevertheless, these findings should be interpreted with caution. Regarding the magnitude of the effect of predictors, low and moderate effect sizes were found. Care must be taken when interpreting them as, although the effect size is intended to represent a "real" effect in the population, it is important to understand that effect size estimates can be influenced by sampling and measurement. Samples that are too large or are non-random, as is the case in this research, may produce biased effect size estimates that should be interpreted with caution (Ferguson, 2009).

This study is not only based on an exploratory analysis of mathematical literacy using machine learning algorithms, but also on a predictive model (hierarchical linear model), taking into account the nested structure of the educational data, in addition to providing the effect size of each predictor. Nonetheless, it should be borne in mind as a limitation that the results do not allow us to conclude in terms of causality and, therefore, it would be interesting to use structural equation models or covariance structures in the future, which are suitable for analysing the type of relationship (direct, indirect or non-causal) and the meaning of the relationships in non-experimental designs.

References

Arroyo-Resino, D., Constante-Amores, A., & Muñoz, I. A. (2019). La repetición de curso a debate: Un estudio empírico a partir de PISA 2015 [Grade repetition under debate: An empirical study from PISA 2015]. *Educación XX1*, *22*(2), 69-92. https://doi.org/10.5944/educxx1.22479

Babore, A., Trumello, C., Candelori, C., Paciello, M., & Cerniglia, L. (2016). Depressive symptoms, self-esteem and perceived parent-child relationship in early adolescence. *Frontiers in Psychology*, *7*, Article 982. https://doi.org/10.3389/fpsyg.2016.00982

Boonk, L., Gijselaers, H. J., Ritzen, H., & Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational Research Review*, *24*, 10-30. https://doi.org/10.1016/j.edurev.2018.02.001

Borgonovi, F., & Pál, J. (2016). A framework for the analysis of student well-being in the pisa 2015 study: Being 15 in 2015. *OECD Education Working Papers*, *1*(40), 1-67. https://doi.org/10.1787/5jlpszwghvvb-en

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Brown, S. (1994). School effectiveness research and the evaluation of schools. *Evaluation and Research in Education*, *8*, 55-68. https:/doi.org/10.1080/09500799409533354

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. https://doi.org/10.1016/j.jrp.2018.02.007

Buzzai, C., Sorrenti, L., Orecchio, S., Marino, D., & Filippello, P. (2020). The relationship between contextual and dispositional variables, well-being and hopelessness in school context. *Frontiers in Psychology*, *11*, 533815. https://doi.org/10.3389/fpsyg.2020.533815

Cameron, A. C., & Windmeijer, A.F. (1997). An R-squared measure of goodness of fit for some Common nonlinear regression models. *Journal of Econometrics*, *77*(2), 329-342. https://doi.org/10.1016/S0304-4076(96)01818-0

Clarke, T. (2020). Children's wellbeing and their academic achievement: The dangerous discourse of 'trade-offs' in education. *Theory and Research in Education*, *18*(3), 263-294.https://doi.org/10.1177/1477878520980197

Cobo-Rendón, C., Pérez, M. V., Romero, H. H., & Parra, J. A. (2017). Modelos multidimensionales del bienestar en contextos de enseñanza-aprendizaje: Una revisión sistemática [Multidimensional models of wellbeing in teaching-learning contexts: A systematic review]. *Ecos de la Academia*, *3*(06), 13-28.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *11*(2), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Constante-Amores, A., Martínez, E. F., Asencio, E. N., & Fernández-Mellizo, M. (2021). Factores asociados al abandono universitario [Factors associated with university dropout]. *Educación XX1*, *24*(1), 17-44. http://doi.org/10.5944/educXX1.26889

Cordero-Ferrera, J. M., Crespo-Cebada, E., & Pedraja-Chaparro, F. (2013). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España [Educational performance and determinants according to PISA: A review of the literature in Spain]. *Revista de Educacion*, *36*(2), 273–97. https://doi.org/10.4438/1988-592X-RE-2011-362-161

Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532-538.

Fiorilli, C., Grimaldi-Capitello, T., Barni, D., Buonomo, I., & Gentile, S. (2019). Predicting adolescent depression: the interrelated roles of self-esteem and interpersonal stressors. *Frontiers in Psychology*, *10*, Article 565. https://doi.org/10.3389/fpsyg.2019.00565

Fleer, M. Rillero, P. (1999). Family involvement in science education: What are the outcomes for parents and students? *Studies in science Education*, *34*(2), 93-114. https://doi.org/10.1080/03057269908560150

Gaviria, J., & Castro, M. (2005). *Modelos Jerárquicos Lineales* [Linear Hierarchical Models].Murralla .

Gil-Madrona, P., Martínez-López, M., & Sáez-Sánchez, M. B. (2019). Objective and subjective factors associated with Spanish students' performance in science in PISA 2015. *Cultura y Educación*, *31*(4), 671-715. https://doi.org/10.1080/11356405.2019.1656485

Gómez-Fernández, N., & Albert, J. (2020). Physical activity in and out-of-school and academic performance in Spain. *Health Education Journal*, *79*(7), 788-801. https://doi.org/10.1177/0017896920929743

Govorova, E., Benítez, I., & Muñiz, J. (2020). How schools affect student well-being: A cross-cultural approach in 35 oecd countries. *Frontiers in Psychology*, *11*, Article 431. https://doi.org/10.3389/fpsyg.2020.00431

Gutiérrez, M., Tomás, J. M., & Pastor, A. M. (2021). Apoyo social de familia, profesorado y amigos, ajuste escolar y bienestar subjetivo en adolescentes peruanos [Social support from family, teachers and friends, school adjustment and subjective well-being in Peruvian adolescents]. *Suma Psicológica*, *28*(1), 17-24. https://doi.org/10.14349/sumapsi.2021.v28.n1.3

Iglesias-Díaz, P., & Romero-Pérez, C. (2021). Aulas afectivas e inclusivas y bienestar adolescente: una revisión sistemática [Caring and inclusive classrooms and adolescent well-being: A systematic review]. *Educación XX1*, *24*(2), 305-350. https://doi.org/10.5944/educxx1.28705

Izquierdo, M.J. (2018). ¿Puede un informe PISA contribuir a una mejor cultura de seguridad y defensa en España? Marco de Competencia Global PISA 2018. [Can a PISA report contribute to a better security and defence culture in Spain? PISA Global Competence Framework 2018]. *Boletín IEEE*, (11), 29-43.

Jorge, E., & González, M. C. (2017). Estilos de crianza parental: Una revisión teórica [Parental parenting styles: A theoretical review]. *Informes Psicológicos*, *17*(2), 39-66. https://doi.org/10.18566/infpsic.v17n2a02

Jungert, T., Levine, S., and Koestner, R. (2020). Examining how parent and teacher enthusiasm influences motivation and achievement in STEM. The Journal of Educational Research, 11 (3), 275–282. https://doi.org/10.1080/00220671.2020.1806015

Kaya, M., & Erdem, C. (2021). Students' well-being and academic achievement: A meta-analysis study. *Child Indicators Research*, *14*(5), 1743-1767. https://doi.org/10.1007/s12187-021-09821-4

Kleinkorres, R., Stang, J., & McElvany, N. (2020). A longitudinal analysis of reciprocal relations between students' well-being and academic achievement. *Journal for Educational Research Online*, *12*(2), 114-165.

Lantz, B. (2013). *Machine learning with R*. Packt.

Laukaityte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics - Theory and Methods*, *46*(22), 11341-11357. https://doi.org/10.1080/03610926.2016.1267764

Lee, V. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, *35*, 125–141. https://doi.org/10.1207/S15326985EP3502_6

López, V., Torres-Vallejos, J., Ascorra, P., González, L., Ortiz, S., & Bilbao, M. (2021). Contributions of individual, family, and school characteristics to Chilean students' social well-being at school. *Frontiers in Psychology*, *12*, Article 620895. https://doi.org/10.3389/fpsyg.2021.620895

López-Martín, E., Expósito-Casas, E., Molina, E. C., & Muñoz, I. A. (2018). ¿Qué nos dice PISA sobre la enseñanza y el aprendizaje de las Ciencias? Una aproximación a través de árboles de decisión [What does PISA tell us about science teaching and learning? A decision tree approach]. *Revista de Educación*, *38*(2), 13-16. 10.4438/1988-592X-RE-

Ministry of Education and Vocational Training. (2019). *PISA 2018. Programa para la evaluación internacional de los estudiantes* [PISA 2018. Programme for International Student Assessment]. Informe español. Secretaría General Técnica.

Molina-Muñoz D, Contreras-García JM and Molina-Portillo E (2023) Does the psychoemotional well-being of Spanish students influence their mathematical literacy? An evidence from PISA 2018. Frontiers in Psychology, 14, 1196529. https://doi.org/10.3389/fpsyg.2023.1196529

Moral-García, J. E., Urchaga-Litago, J. D., Ramos-Morcillo, A. J., & Maneiro, R. (2020). Relationship of parental support on healthy habits, school motivations and academic performance in adolescents. *International Journal of Environmental Research and Public Health*, *17*(3), Article 882. https://doi.org/10.3390/ijerph17030882

Murillo, F. J., & Hernández, R. (2020). ¿La implicación de las familias influye en el rendimiento? Un estudio en educación primaria en América Latina [Does family involvement influence achievement? A study in primary education in Latin America]. *Revista de Psicodidáctica*, *25*(1), 13-22. https://doi.org/10.1016/j.psicod.2019.10.002

OECD. 2009. PISA 2006 *Technical Report*. OECD Publishing. https://www.oecd.org/pisa/data/42025182

OECD. 2017. *PISA 2015 Results (Volume III): Students' Well-Being*. OECD Publishing. https://dx.doi.org/10.1787/9789264273856-en

OECD. 2019. *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. https://dx.doi.org/10.1787/b25efab8-en

Pardo, M., & Ruiz, M. (2013). *Análisis de Datos en Ciencias Sociales y de la Salud III* [Data Analysis in Social and Health Sciences III]. Síntesis.

Park, S., Stone, S. I., & Holloway, S. D. (2017). School-based parental involvement as a predictor of achievement and school learning environment: An elementary school-level analysis. *Children and Youth Services Review*, *8*(2), 195-206. https://doi.org/10.1016/j.childyouth.2017.09.012

Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python*. Packt.

Raudenbush, S. W., & Willms, J. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*(1), 307-335. https://doi.org/10.3102/10769986020004307

Rivero, G. (2011). *Análisis de datos incompletos en ciencias sociales* [Incomplete data analysis in social sciences]. CIS

Rodríguez, A. D., Rivas, J., & Adasme, G. P. (2020). Tiempo de compromiso motor en la clase de educación física [Motor engagement time in the physical education class]. *Voces de la Educación*, *5*(10), 90-113. https://doi.org/10.25053/redufor.v5i14mai/ago.2895

Sánchez, E. M., Miguélañez, S. O., & Abad, F. M. (2019). Explanatory factors as predictors of academic achievement in PISA tests. An analysis of the moderating effect of gender. *International Journal of Educational Research*, *96*, 111-119. https://doi.org/10.1016/j.ijer.2019.06.002

Sheridan, S. M., Smith, T. E., Moorman Kim, E., Beretvas, S. N., & Park, S. (2019). A meta-analysis of family-school interventions and children's social-emotional functioning: Moderators and components of efficacy. *Review of Educational Research*, *89*(2), 296-332. https://doi.org/10.3102/0034654318825437

Shirley, D., Hargreaves, A., & Washington-Wangia, S. (2020). The sustainability and unsustainability of teachers' and leaders' well-being. *Teaching and Teacher Education*, *92*, 1-12. https://doi.org/10.1016/j.tate.2019.102987

Sjøberg, S. & Jenkins, E. (2022). PISA: A political project and a research agenda. *Studies in Science Education*, *58*(1), 1-14. https://doi.org/10.1080/03057267.2020.1824473

Smith, T. E., Sheridan, S. M., Kim, E. M., Park, S., & Beretvas, S. N. (2020). The effects of family-school partnership interventions on academic and social-emotional functioning: A meta-analysis exploring what works for whom. *Educational Psychology Review*, *32*(2), 511-544. https://doi.org/10.1007/s10648-019-09509-w

Tan, C. Y., Lyu, M., & Peng, B. (2020). Academic benefits from parental involvement are stratified by parental socioeconomic status: A meta-analysis. *Parenting*, *20*(4), 241-287. https://doi.org/10.1080/15295192.2019.1694836

Tomé, G., Almeida, A., Ramiro, L., & Gaspar, T. (2021). Intervention in Schools promoting mental health and well-being: A systematic review. *Global Journal of Community Psychology Practice*, *12*(1), 1-23.

Trigueros, R., Aguilar-Parra, J. M., Cangas, A. J., Bermejo, R., Ferrandiz, C., & López-Liria, R. (2019). Influence of emotional intelligence, motivation and resilience on academic performance and the adoption of healthy lifestyle habits among adolescents. *International Journal of Environmental Research and Public Health*, *16*(16), 1-12. https://doi.org/10.3390/ijerph16162810

Urrila, A. S., Artiges, E., Massicotte, J., Miranda, R., Vulser, H., Bézivin-Frere, P., Lapidaire, W., Lemaître, H., Penttilä1, J., Conrod, P.J., Garavan, H., Paillère- Martinot, M.L., Martinot, J.L., & the IMAGEN consortiumet. (2017). Sleep habits, academic performance, and the adolescent brain structure. *Scientific Reports*, *7*(1), 1-9. https://doi.org/10.1038/srep41678

Wang, Y., Huebner, E. S., & Tian, L. (2021). Parent-child cohesion, self-esteem, and academic achievement: The longitudinal relations among elementary school students. *Learning and Instruction, 73,* Article 101467. https://doi.org/10.1016/j.learninstruc.2021.101467

Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors predicting mathematics achievement in PISA: a systematic review. *Large-scale Assessments in Education*, *11*, 24. https://doi.org/10.1186/s40536-023-00174-8

Widlund, A., Tuominen, H., & Korhonen, J. (2018). Academic well-being, mathematics performance, and educational aspirations in lower secondary education: Changes within a school year. Frontiers in psychology, 9, 297. https://doi.org/10.3389/fpsyg.2018.00297

Wu, X., Gai, X., & Wang, W. (2020). Subjective well-being and academic performance among middle schoolers: A two-wave longitudinal study. *Journal of Adolescence, 84*, 11-22. https://doi.org/10.1016/j.adolescence.2020.07.011

Yang, Q., Tian, L., Huebner, E. S., & Zhu, X. (2019). Relations among academic achievement, self-esteem, and subjective well-being in school among elementary school students: A longitudinal mediation model. *School Psychology*, *34*(3), 32-38. https://doi.org/10.1037/spq0000292