IJPES
www.ijpes.com
International Journal of Psychology and Educational Studies
ISSN: 2148-9378

Investigation of Turkish Students' School Engagement through Random Forest Methods Applied to TIMSS 2019: A Problem of School Psychology

Hikmet ŞEVGİN[1], Anıl Kadir ERANIL[2]

[1] *Faculty of Education, Van Yüzüncü Yıl University, Van, Türkiye* *0000-0002-9727-5865*

[2] *Nevşehir National Education Directorate, Nevşehir, Türkiye* *0000-0001-7804-735X*

ARTICLE INFO

Article History
Received 18.06.2023
Received in revised form 13.09.2023
Accepted 06.10.2023
Article Type: Research Article

ABSTRACT

In this study, in line with TIMSS 2019 data, 8th grade students' school engagement in the Turkish education system and various variables in the field of science and mathematics, which are thought to be related to school engagement, were examined with the Random Forest method, a method amongst data mining methods. This research focuses on student bullying, which includes psychological factors at school, such as whether students like or dislike their lessons, their confidence in science, their self-confidence, absenteeism, and the effects of teachers' teaching methods on SE. The sample of the study consisted of 3872 students in the science data set and 3802 students in the mathematics data set, which remained as a result of the lost data deletion and assignment processes from 4077 students who originally participated in the application. The open-source Python infrastructure was used in the analysis of the data. Orange 3.32 data mining program was employed for model setup. The model performance criteria MSE, RMSE, MAE and R2 values obtained as a result of the analysis. In both areas, the variables that contribute to the prediction of students' school engagement were ranked according to their importance levels, starting from the most important, also interpreted and discussed. It was observed that the performance criteria of the established model have values close to zero in the field of science (MSE: 2.775 RMSE: 1.666 MAE: 1.267) and mathematics (MSE: 2.240 RMSE: 1.497 MAE: 1.131). Variables explain school engagement at the rate of 69.6% in science and 75.7% in mathematics. The order of importance of the variables in both areas showed a great similarity. Student bullying was obtained as the most important variable. Prospective studies can also be planned towards collecting more in-depth data deploying qualitative data collection methods under a qualitative research model that also includes the opinions of self, peers, teachers and parents. For the education policies that the TES should produce toward increasing SE, it is weighty to reduce and prevent student bullying in schools with sustainable practices. Exclusively school administrations should focus on student bullying with the help of counselors, diagnose the problems and take measures in and out of school according to the types of bullying.

Keywords:
School psychology, school engagement, data mining, random forest

1. Introduction

Ensuring that students are educated as individuals who are well-equipped in all aspects of attaining academic success and establishing healthy social relations is an output desired by all education systems. First of all, students need to love school and thusly wish to spend time in the school environment. Arguably, education systems can achieve their goals in this way. To that end, it would be fair to say that school engagement (SE) has a positive effect on students, both in terms of creating and sustaining academic and social relations as desired. Similarly, SE is considered a solution to problems emerging on the way to reaching the goals, as it is closely related to increasing students' academic achievement and improving their learning skills (Ladd &

[1] Corresponding author's address: Van Yüzüncü Yıl University, Faculty of Education, Van /Türkiye
e-mail: hikmetsevgin@gmail.com

Citation: Şevgin, H. & Eranıl, A. K. (2023). Investigation of Turkish students' school engagement through random forest methods applied to TIMSS 2019: A problem of school psychology. *International Journal of Psychology and Educational Studies*, *10*(4), 896-909. https://dx.doi.org/10.52380/ijpes.2023.10.4.1260

Dinella, 2009). In addition, students with high SE have more positive social relationships with their schoolmates and have higher social competencies (Demirtas-Zorbaz et al., 2018). Therefore, SE is deemed pivotal by researchers as it increases both the social and academic development of the student (O'Farrell et al., 2006).

Although not wanting to be in the school environment may arise as a student-related reason, it is oftentimes caused by obligations such as the environment or family conditions of the student. Thus, SE should not be understood as a phenomenon that is solely student-dependent. Family, peers, and the school itself are the three main contexts associated with SE (Fernández-Zabala et al., 2016). SE, which gains a multifaceted and intricate structure, then concerns each of the education stakeholders in this framework.

All these situations are also present in the Turkish education system (TES). When the research results in TES are examined, it is evident that SE is a subject that needs to be investigated with various variables. Moreover, it is understood that school is the place where a child should be in terms of healthy development. Therefore, it is necessary to determine each factor that increases the child's commitment to school. The fact that SE has a place in the education system in practice can be considered a way to not only educate students with high academic success but also to educate children in schools as healthy individuals in all aspects.

There is also a relationship between academic success and SE. According to the results of Kaya and Boyraz's (2020) meta-analysis research, SE has a weak positive effect on students' academic achievement. According to the results of Özdemir and Yalçın's (2009) research, there is a positive correlation between students' SE levels and their Turkish, basic mathematics, social sciences, and science scores. According to Altuntaş and Sezer's (2017) research results, loss of interest in school and burnout due to school activities significantly predict SE. They found a significant relationship between the variables of grade level, academic achievement, physical structure of the school, teacher satisfaction, number of friends, and SE.

In line with all these thoughts and research, SE has a key role in education in all aspects. This research aims to determine the variables that explain the science and mathematics scores of the 8th grade students in TES in line with the TIMSS 2019 Turkey data and to examine the effects of these variables on the SE.

SE is the broadest and most frequently used comprehensive meta-structure that analyzes students' relationships with their schools (Fredricks et al., 2004) and is defined as the frequency of students' participation in academic and non-academic school activities associated with school outcomes (OECD, 2003). Along similar lines, SE has multiple dimensions and multiple contexts due to its encompassing nature (O'Farrell et al., 2006).

SE is divided into three major themes, viz., behavioral, emotional, and cognitive engagement (Fredricks et al., 2004; Jimerson et al., 2003; Li et al., 2020). Further, according to Li, Gao, and Sha (2020), SE refers to the behavioral, emotional, and cognitive engagement of students in learning activities and the intensity of the emotions they experience during the said processes. Cognitive school engagement is intellectual progress that is associated with the difficulties students experience in the classroom, the activities they perform, and the efforts they make to achieve their learning goals (Ay, 2022). When all this technical structure is taken together, the three dimensions of SE in question are observable outward expressions, and the emotional reactions of a learner to an activity or assignment can be observed in this direction through their verbal or nonverbal expressions (Erentaitė et al., 2018).

A school should have a number of attractive features for students and for professional educators so that dropouts decrease and SE increases. According to Kukuş and Akto (2022), students want to be educated in schools that are well-equipped physically and where a positive rapport is created from within, especially with friends and teachers of students, which yields positive feelings towards school. A similar situation remains valid for TES as well. According to the results of Özbilen, Eranıl, and Özcan's (2018) research, students who are successful in their courses are also more committed to school than those who are not successful, and as students' relationships with their families and friends increase, so does their commitment to school.

SE is an eminent matter that the TES tries to assure but also carries various difficulties in this regard. This is echoed in the line of literature too. Arastaman (2009) declared that students' commitment to school is exam-oriented, and students show commitment to school in order to pass their courses. The notion of SE emerges as one that needs to be clarified for TES in this sense. Furthermore, it holds strategic importance in that as the grade levels of the students go higher in TES, their commitment to school declines (Şahan & Özgenel, 2021) and in that the students who come to school by bus, comprising a disadvantaged group, have higher levels of SE than those who do not have transportation (Uzun, 2021). Vocational high schools can also be counted as a disadvantaged group because students with low academic success continue their education at vocational high schools in order to acquire professional skills. It can be put forth here that these students do not have a bright academic background, particularly when thinking about their primary and secondary school lives. Taking these into consideration, SE in TES appears to carry a prominent role, in particular in terms of enabling disadvantaged students to carry on education in a formal environment. Apart from these, Ünlü, Evcin, Yılmaz, and Dalkılıç (2013) put forward that crime and violence tendencies in vocational high schools are more prevalent than in other high school types. Vis-à-vis SE, this can be further supported by the study of Özbilen, Eranıl, and Özcan (2018), who announced that there is a negative correlation between students' SE and aggression in TES. In TES, a negative correlation was found between SE, deviant behaviors, and substance use (Ünal & Çukur, 2011).

TIMSS is an international test and calculates possible values for each student. Because of its focus on TIMSS data, this research focused on the mathematics and scores of 8th grade students in Turkey on the axis of school engagement. In summary, the Turkish results of the TIMSS data presenting some findings on the commitment of 8th graders to school may shed light on how to ensure SE in TES. Although some studies have been conducted on SE as a result of the research mentioned above, no study has been found that comprehensively investigates students' SE in the Turkish education system through the Random Forest (RF) method within the scope of educational data mining in line with the data obtained in the field of education. The main problem situation that this research focuses on is 'What are the various variables that explain the level of school engagement of 8th grade students in the Turkish education system through their mathematics and science achievement scores?' In this context, answers to the following questions are sought:

- Based on Random Forest data mining method in line with TIMSS 2019 Turkey science scores, SE's'student bullying, student values science, students like learning science, instructional clarity in science lessons, student confidence in science, absenteeism, intended graduation level in education', what are the significance levels in terms of predictors?
- Based on Random Forest data mining method in line with TIMSS 2019 Turkey mathematics scores, SE's'student bullying, instructional clarity in mathematic lessons, students like learning mathematics, students value mathematics, students confident in mathematics, disorderly behavior in mathematics during lessons, absenteeism, intended graduation level in education', what are the significance levels in terms of predictors?

2. Methodology

2.1.Research Model

This research adopts a quantitative paradigm and was designed using a correllation survey model. Using correllation surveying, information about a large group can be received by examining a sample (Leedy & Ormrod, 2005). The correllation survey model refers to the correlational and regressional approach between the predicted variable and the predictive variables.

2.2. Study Group

The data for this study consists of 4077 Turkish students who participated in the TIMSS 2019 application. Before starting the analyses, the deletion of missing data in demographic data was performed. In this process, students who left the demographic variables unanswered, about which information was tried to be collected with a single answer, were excluded from the data set. This process was applied to six variables, 3 in the field of science and three in the field of mathematics. Thus, data belonging to 205 (5%) students in science and 275 (6.5%) students in mathematics were removed from the data set by applying the deletion process via index. As a matter of fact, in this study, it can be stated that the change in the standard deviation and the power of the test were affected at a low level because the sample size was large enough and the deletion process was low. Due to the fact that demographic data includes deletion and data obtained at the Likert-type scale level, MCAR (totally unbiased loss) missing data for each scale is less than 5% in total (Tabachnick & Fidell, 2015), value assignment according to the mean was applied. In this process, in order not to reduce the variance and correlation of the variables, the averages of the items were taken into account on the basis of each item, not on the basis of the variable. In Annex-1, information on the first and last mean and standard deviation values on the basis of each item and how many cells were assigned to each item is given in tabular form. Thus, 3872 students in the science data set and 3802 students in the mathematics data set constitute the sample of this study.

2.3. Data Collection Tools and Procedure

The TIMSS 2019 application contains several cognitive and affective scales prepared at the scale level in the field of science and mathematics and items composed of demographic information about students. In this study, science and mathematics fields were inspected separately. What is more, the data belonging to the variables determined specifically for these areas on SE was delved into through the 'Random Forest' data mining method.

In the model established in the field of science, SE (5 items) was the dependent variable, student bullying (14 items), student values science (9 items), instructional clarity in science lessons (7 items), students like learning science (9 items), and students confident in science (8 items) were included in the model as independent variables. In the model established in the field of mathematics, SE (5 items) was the dependent variable, while student bullying (14 items), students value mathematics (9 items), instructional clarity in math lessons (7 items), students like learning mathematics (9 items), students confident in mathematics (9 items), and disorderly behavior during mathematics lessons (6 items) were included in the model as independent variables. In addition, the categorical variables of absenteeism and intended graduation level in education were also included in both science and mathematics models. All variables obtained at the scale level are of the 4-point Likert type. In the determination of these variables, the opinions of the experts were taken into account. Opinions of three experts in each field, including two field experts and one measurement and evaluation expert from each of the fields of science and mathematics, and a total of five experts were taken. The voting technique was used in the determination phase, and the variables that received at least two out of three votes were included in the model.

In order to provide evidence for the validity of the scales, the results of the exploratory factor analysis specified in the TIMSS 2019 technical report were examined. Each scale has a single factor, and the lowest item is 0.51 while the highest is 0.89 (Martin et al., 2020). It is understood that the relevant features of the items belonging to the scales are good representatives. Reliability indices (Chronbach alpha) presented at the scale level were also examined. It was observed that only the belonging to school scale (0.76) was below 0.80, and all scales had a reliability index above 0.80. In addition, the highest reliability was found to be 0.92 on the love of learning mathematics scale (Martin et al., 2020). In this study, the reliability index of the scales was recalculated due to data deletion and assignment operations. It was observed that the same values were obtained, with the lowest 0.76 belonging to the school belonging scale and the highest 0.92 for the love of learning mathematics scale. In the related literature, $0.60 \leq \alpha \leq 0.80$ is considered highly reliable, and $0.80 \leq \alpha \leq 1.00$ is highly reliable (Özdamar, 2013). Reliability values arrived at can be expressed as values above the accepted limit values.

The histogram and Q-Q graphs were examined for the univariate normal distribution of the dependent variable SE. Evidence was also collected through skewness-kurtosis values and hypothesis testing. The fact that the normal distribution curve drawn on the histogram graph did not comply with the graph, while the points on the Q-Q graph showed direct deviations from the plot, showed a deviation from normality. The fact that the skewness (-1.324) and kurtosis (-1.728) values are outside the range of -1 to +1 indicates a deviation from normality. The fact that the p value obtained as a result of the Kolmogorov-Smirnov test, which is one of the hypothesis tests for normality, is lower than 0.05 indicates a deviation from normality. All this evidence reveals that the dependent variable does not exhibit a normal distribution. For multivariate normality, univariate normality must be provided. Therefore, since univariate normality could not be achieved, multivariate normality was not examined. In the graph showing the standardized residual values and the estimated values for linearity, it was observed that there was no symmetrical pattern above and below the y=0 line, but a curved pattern, and there was no linearity between the variables. In addition, the homogeneity of variances was tested with Levene's test, and the p value was found to be less than 0.05. Therefore, it was concluded that the variances were not homogeneous. Findings related to all these assumptions show that non-parametric tests should be preferred rather than parametric tests such as multiple linear regression. In this context, the TIMSS 2019 Turkey dataset was analyzed using the 'Random Forest' method, which is a non-parametric test type.

Prior to the analysis, it was tested to see if there was a multicollinearity problem among the variables included in the analysis. In multicollinearity tests, Variance Inflation Factor (VIF) and Tolerance values of multicollinearity are allowed. Sould the VIF value is greater than 10 or the Tolerance value is less than 0.1, it is evident that there is a multicollinearity problem between the variables (Keller et al., 2012). In this study, it was perceived that the VIF values for the Science data set ranged from 1.042 to 2.537, the Tolerance values ranged between 0.394 and 0.959, the VIF values for the mathematics data set ranged between 1.144 and 2.586, and the Tolerance values ranged between 0.387 and 0.874. Accordingly, there is no multicollinearity problem between the variables used in the research.

2.4. Data Analysis

As in all fields, in the field of education, the preferred analysis method also changes depending on whether the data set is parametric or non-parametric. Recently, data mining and machine learning methods have been preferred in the analysis of all data sets, irrespective of whether they are parametric or not (Strobl et al., 2009). It can be said that it ensures richness in the variety of analysis that can be employed in the analysis of data sets demonstrating a non-parametric structure (Strobl et al., 2009). The random forest (RF) data mining method is an ensemble learning method. RF trains many classification and regression trees (C&RT) from the family of decision trees, randomly using the bagging method to obtain a more accurate prediction among them, and combines multiple C&RT trees to form a forest. This created forest is a C&RT community trained with the bagging method. Bagging constitutes the final classifier of the models for classification through the multiple voting method and the final estimator for the regression models with the average of the parameter estimates (Ferreira & Figueiredo, 2012). Therefore, the RF method is used in both classification and regression analysis (Breiman, 2001).

RF consists of two stages. In the first stage, it creates a random forest from decision trees (C&RT, etc.). It then makes predictions on the generated random forest classifier. It is known to be used in many fields, such as econometrics, medicine, banking engineering, and education.

That said, some 'tunning' settings need to be made during the model setup phase. First off, the value equal to the square root of the number of attributes (dependent and independent variables) in the data should be chosen for the number of variables to be randomly selected at each node (Breiman Last & Rice, 2003). Since the number of features in the model established for the science field is 8, the square root is approximately 2.82. As it is necessary to enter an integer, this number is selected as 3 by completing one on top of the other. Correspondingly, as the number of features in the model established for the field of mathematics is 9, this number has been chosen as 3 with its square root. The next step is to decide on the number of trees sufficient for analysis. In this case, it is significant to appoint the most suitable number of trees to form the decision forest. For this, the number of trees with the lowest error values indicates a sufficient number of trees (Huffer & Park, 2020; Probst & Boulesteix, 2017). The RF method does not need to provide the assumptions that the data set needs to be provided for the parametric methods to be preferred. In this respect, the RF data mining method is a non-parametric method that does not require assumptions such as normality, linearity, and homogeneity (Geneur et al., 2017; Biau & Scornet, 2016).

In this research, the 'Random Forest' analysis method, which is one of the data mining methods, was selected because the data set exhibits non-parametric properties pertaining to its structure. The open-source Python infrastructure was referred to in the analysis of the data, and the Orange 3.32 data mining program was utilized for model setup.

Model Performance Metrics

MSE (Mean Squared Error) $$\boldsymbol{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

In general terms, mean square error refers to the closeness of a set of points to the regression curve MSE measures the amount of error of an estimator, always has a positive value, and evaluates the mean square difference between observed and predicted values. Estimators with an MSE close to zero perform better (Bickel vd., 2015).

RMSE (Root Mean Square Error) $RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$

It is equal to the square root of the mean square error value. It is a metric that assesses the magnitude of error and is often used to find the distance between an estimator's predicted values and their true values. The RMSE value can range from 0 to ∞. Closer to zero signals that the error has diminished. If the RMSE value is zero, it means that the model has not made any errors (Willmott & Matsuura, 2005).

MAE (Mean Absolute Error) $MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$

The mean absolute error is a measure of the errors between two continuous variables, the predicted value and the observed value. MAE is the average vertical distance between each true value and the line that best fits the data. MAE is also the average horizontal distance between each data point and the best-fit line. The MAE value can range from 0 to ∞. Estimators with an MAE close to zero perform better (Willmott & Matsuura, 2005).

R^2 (Coefficient of Determination) $R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\hat{y} - y_i)^2}$

The coefficient of determination (R^2) is a statistical measure that represents the ratio of the variance explained by the independent variables in the regression model for the dependent variable (Glantz vd., 1990).

2.5. Ethical

The research was carried out with the approval of the Van Yüzüncü Yıl University Ethics Commission, dated July 25, 2022, and numbered 2022/16-03.

3. Findings

In this section, the findings obtained in terms of metric values are shared in order, first for the science field and then for the mathematics field.

Science Score and SE: For the model established in the field of science, the most suitable number of trees to form the decision forest has been determined. The analysis of tree numbers was repeated for each selection as 100, 250, 500, 750, 1000, 1500, 2000, and 3000. The results obtained are presented in Table 1.

Table 1. *Determination of the Maximum Sufficient Number of Trees According to the Metrics Obtained in Science*

Field	Number of trees	MSE	RMSE	MAE	R^2
Science	100	2.811	1.676	1.273	0.692
	250	2.790	1.670	1.272	0.695
	500	2.787	1.670	1.270	0.695
	750	2.775	1.666	1.267	0.696
	1000	2.776	1.668	1.269	0.696
	1500	2.776	1.666	1.267	0.696
	2000	2.776	1.667	1.268	0.696
	3000	2.778	1.667	1.268	0.696

As can be seen in Table 1, the lowest error value was obtained for the number of 750 trees. In all subsequent tree numbers, it was observed that the error rate was always the same or close to it. It can be said that the same or close values are obtained no matter how much the number of trees increases. The analysis results obtained for the model established in this study were obtained with a total of 750 trees.

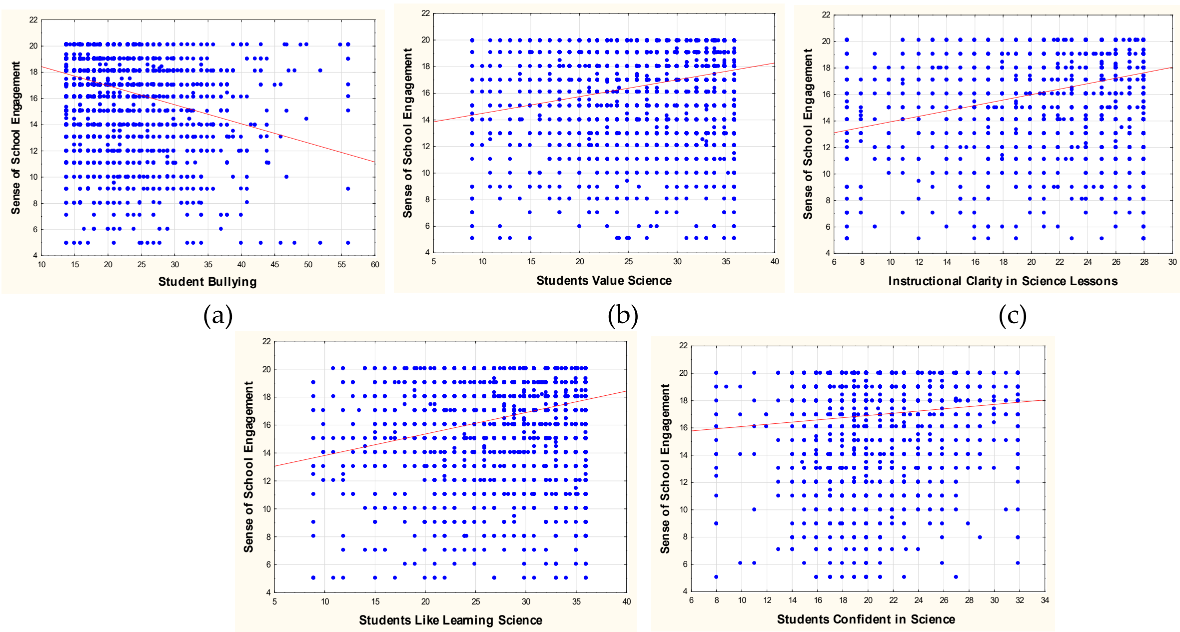
The MSE value is 2.775, the RMSE value is 1.666, the MAE value is 1.267, and R^2 was obtained as 0.696. Evidently, the MSE, RMSE, and MAE values are close to zero. The fact that these values are zero means that the established model does not have any mistakes. In this direction, the obtained values indicate that the error rate is low. It can be said that the values estimated by the method are close to the real values. The percentile of the coefficient of determination reveals the explained variance of the model. In this study, it was realized that the independent variables included in the model explain 69.6% of the dependent variable. The most important predictors and significance levels of these variables on the dependent variable in the established model are given in Table 2.

Table 2. *Significance Levels of Variables in Science According to the Random Forest Method*

	Variables	Data type	Severity level	%
1	Student bullying	Continuous	0.232	23,3
2	Students value science	Continuous	0.162	16,2
3	Students like learning science	Continuous	0.161	16,1
4	Instructional clarity in science lessons	Continuous	0.155	15,5
5	Student confident in science	Continuous	0.145	14,5
6	Absenteeism	Categorical (5)	0.023	2,3
7	Intended graduation level in education	Categorical (6)	0.018	1,8

Categorical 5: "Hardly ever" Categorical 6: "Postgraduate"

When Table 2 is viewed, the most important predictors of SE are: student bullying, student values of science, students like learning science, instructional clarity in science lessons, student confidence in science, absenteeism, and intended graduation level in education. The scatter plots showing the relationship between the variables that are continuous among the independent variables and SE are presented in Figure 1.



(a) (b) (c)

(d) (e)

Figure 1. *Scatter Plots Displaying the Relationships Between the İndependent Variables in Science and the Dependent Variable (SE)*

As can be seen in Figure 1a, there is a negative relationship between student bullying and SE. SE decreases as long as student bullying increases. In Figure 1b, SE increases as student values of science increase. In Figure 1c, the SE increases as the instructional clarity in science lessons increases. Similarly, SE increases as students like learning science increase (Figure 1d). Finally, the SE increases as student confidence in science increases, albeit at a low level.

Mathematic Score and SE: For the model established in the field of mathematics, the most appropriate number of trees has been determined to create the decision forest. In this frame, the analyses were repeated for each selection, with tree numbers of 100, 250, 500, 750, 1000, 1500, 2000, and 3000. The results reached are presented in Table 3.

As can be accessed through Table 3, the lowest error value was obtained at the number of 1000 trees. In all subsequent tree numbers, it was unearthed that the error rate was always the same or close to it. This can be accounted for by the fact that the same or close values are obtained no matter how much the number of trees rises. The analysis results obtained for the model established in this study were obtained with a number of 1000 trees.

Table 3. *Determination of the Maximum Sufficient Number of Trees According to the Metrics Obtained in the Field of Mathematics*

Field	Number of trees	MSE	RMSE	MAE	R^2
Mathematics	100	2.300	1.517	1.143	0.750
	250	2.251	1.500	1.135	0.756
	500	2.266	1.505	1.136	0.754
	750	2.250	1.500	1.132	0.756
	1000	2.240	1.497	1.131	0.757
	1500	2.240	1.500	1.133	0.756
	2000	2.241	1.497	1.131	0.757
	3000	2.242	1.499	1.132	0.756

The MSE value is 2.240, the RMSE value is 1.497, the MAE value is 1.131, and R^2 was obtained as 0.757. It is noticed that the MSE, RMSE, and MAE values are close to zero, so one may say that the obtained values have a low error rate. As a result, the model verifies that the values estimated by the method are close to the true values. On top of these, it is contemplated that the independent variables included in the model explain 75.7% of the dependent variable. The most important predictors and significance levels of these variables on the dependent variable in the established model are communicated in Table 4.

Table 4. *Significance levels of variables in the field of mathematics according to the random forest method*

	Variables	**Data type**	**Severity level**	**%**
1	Student bullying	Continuous	0.173	17,3
2	Instructional clarity in math lessons	Continuous	0.170	17,0
3	Students like learning mathematics	Continuous	0.166	16,6
4	Students value mathematics	Continuous	0.127	12,7
5	Student confident in mathematics	Continuous	0.123	12,3
6	Disorderly behavior during mathematics lessons	Continuous	0.118	11,8
7	Absenteeism	Categorical (5)	0.018	1,8
8	Intended graduation level in education	Categorical (6)	0.015	1,5

When Table 4 is scanned, the essential predictors of SE are, respectively, flows: student bullying, instructional clarity in mathematic lessons, students like learning mathematics, students value mathematics, students are confident in mathematics, disorderly behavior during lessons, absenteeism, and intended graduation level in education are variables. Figure 2 illustrates the scatter plots showing the relationships between the continuous variables, which are independent variables, and the SE.

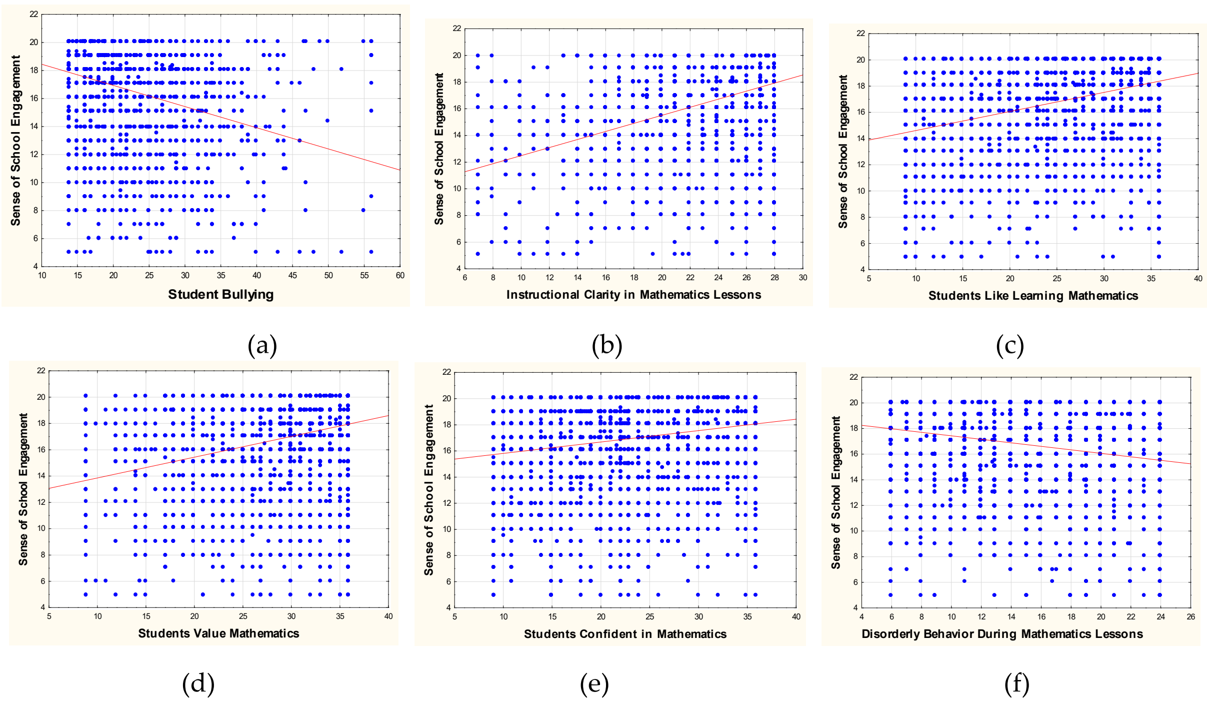


Figure 2. *Scatter Plots Showing the Relationships Between Independent Variables in Mathematics and the Dependent Variable (SE)*

As can be seen in Figure 2a, SE goes down as long as student bullying goes up. Put differently, there is a negative relationship between them. In Figure 2b, it was noted that SE increased with intensified instructional clarity in math lessons. In figure 2c, students' liking for learning mathematics increases as the SE multiplies. Uniformly, in figure 2d, SE augments as the students' mathematical value progresses. The SE of students confident in mathematics heightens, albeit to a small extent. Finally, there is a negative relationship between disorderly behavior during mathematics lessons and SE. It was noted that the SE decreased as the disorderly behavior during mathematics lessons increased.

4. Conclusion and Discussion

Within the scope of this study, the SE of 8th grade students participating in the TIMSS 2019 application from Turkey and various variables in the field of science and mathematics, which are included in the application and are thought to be related to SE, were examined using the Random Forest method, a data mining method. Variables related to school attachment (SE) were included in the model separately for the science and mathematics fields, and analyses were executed. In this regard, the relationship between various variables that explain the level of SE of 8th grade students in the TES in the fields of mathematics and science was dwelled upon. Moreover, the ways in which these variables serve SE or dropout were paid attention to in light of the available data. It might be concluded that each variable included in the model significantly contributes to the prediction of students' commitment to school (SE).

The order of the significance levels obtained as a result of the analysis of the variables included in the model in science and mathematics is quite similar. In both, student bullying is obtained as the most important variable. In other words, according to the first result of the research,'student bullying' is ranked at the top in both fields, and it is the variable that contributed the most to predicting students' SE. Reducing student bullying significantly amplifies students' SE. SE is a complex development process (Perdue et al., 2009). Knowing that, it may be argued that a student's SE is also affected by numerous situations, such as the school environment, friend relations, teacher relations, and family structure. Family support and student-family relations are among the most important factors that impact students' SE (Atoum et al., 2019; Erol & Turhan, 2018; Mehta et al., 2013; Mo & Singh, 2008; Murray, 2009; Revuelta et al., 2018).

Li, Chen, and Li (2020) emphasized in their study that it is paramount to develop students' sense of SE against peer bullying because, according to Yang (2015), one of the most negative consequences of student bullying is school dropout Then, the development of the sense of SE, which is a complex structure, attracts the attention of educators (Li et al., 2020; Luo et al., 2020; Önen, 2014). This hints at the fact that, on account of student bullying in the school environment, the student does not feel comfortable or safe at school, and their commitment to school degrades. Consonantly, Özdemir and Yalçın's (2019) study recorded a positive relationship between students' perceptions of school safety and their Turkish, mathematics, and science scores. According to Aulia (2016), when students are bullied at school, they develop negative emotions and are exposed to physical and verbal bullying in their relationships with their peers. How the child feels at school emerges as an exigent parameter in this context. According to Reschly, Huebner, Appleton, and Antaramian (2008), children who feel positive at school are more willing to learn, while those who feel negative develop less SE. Alongside these, children who are involved in bullying behaviors are also prone to harmful behaviors such as drinking alcohol and smoking (Cosma et al., 2015).

The quality of friendship and peer support of a child with peers advance SE (Perdue et al., 2009). Along with this, teacher bullying and peer bullying significantly influence student behavior (Najam & Kashif, 2018). Peer social support can be considered the opposite of peer bullying, and peer social support affects students' school adjustment more strongly than teacher social support (Wang & Eccles, 2012). Similarly, Mehta, Cornell, Fan, and Gregory (2013) underpinned in their study that peer bullying is effective on students' school participation. Han, Kang, Choe, and Kim (2021) uncovered a negative relationship between peer bullying and students' SE feelings in the study they conducted with students from the Northeast Asian region and the United States who participated in the PISA 2015 application. It can be deduced that the results of this research and the literature information are similar. Lowering peer bullying both adds to the SE of students, resulting in contributions to their academic development. In like manner, according to the results of the research, there is a positive relationship between SE and students' academic desire and development (Veiga et al., 2014).

The quality of a child's friendship with peers and peer support are influential factors in school engagement (Perdue et al., 2009). Teacher bullying as well as peer bullying significantly impact student behavioral participation (Najam & Kashif, 2018). Peer social support can be considered the opposite of peer bullying, and peer social support influences students' school adjustment more strongly than teacher social support (Wang & Eccles, 2012). Along similar lines, Mehta, Cornell, Fan, and Gregory (2013) articulated that peer bullying has an effect on students' school participation.

Han, Kang, Choe, and Kim (2021), in their study on students from Northeast Asia and the United States who participated in the PISA 2015 application, documented that peer bullying had a negative relationship with students' sense of school belonging. It can be asserted that the results of this study and the results in the literature are similar. Lessening peer bullying proliferates students' commitment to school and, as a consequence, paves the way for academic development. Consistently, according to the results of the research, there is a positive relationship between school engagement and students' academic desire and development (Veiga et al., 2014).

According to the second series of results of this research,'student values lesson' is ranked second in science and fourth in mathematics among the variables contributing to students' SE prediction. The fact that students' levels of valuing lessons are high also strengthens the feeling of SE. Together with this,'students like learning lessons' is ranked third in both fields (science and mathematics) according to their level of importance. Similar to this result, Smith, Walker, Chen, and Hong (2020), in their study on the data from the TIMSS 2015, came to know that students' love of science and mathematics was a significant and positive predictor of SE. Japelj Pavešić, Radović, and Brese (2022) enunciated in their study on the data obtained from the countries in the Dinaric region participating in the TIMSS 2019 that the students with a high sense of SE also had a high love for science and mathematics lessons. It can be proffered that the student's liking and valuing of the lesson increases the SE. Love and value can also be strengthened by the student's positive perceptions of the school environment. According to Wang and Eccles (2013), students' perceptions of the school environment increase their motivation for success. By the same token, students' emotional commitment to school is also predicted by the teacher's trust in parents and students (Ahmed et al., 2021). According to the results of Gunuc's (2014) research, it was recognized that lesson engagement predicts academic success and explains it by 10%.

According to the final group of results of this study, 'instructional clarity in lessons' is ranked fourth in science and second in mathematics among the variables contributing to students' SE prediction. 'Student confident' ranked fifth in both fields according to the level of importance.

Beyond this, unlike the science field, the 'disorderly behavior during mathematics lessons' obtained in the field of mathematics is ranked sixth in the field according to the level of importance. In view of this, when the literature is searched, SE and school burnout are momentous variables that affect students' personal, academic, and social adaptation (Demirci et al., 2020). In the study of Fernández-Zabala, Goñi, Camino, and Zulaika (2016), the teacher's support for the student was also found to have the strongest correlation value. On this basis, the student's lack of self-confidence can also be interpreted as a behavioral disorder and may be negatively affecting SE. According to Atilola, Abiri, and Ola (2022), behavioral disorders have a negative impact on young people's participation in school and incorporate a potential barrier to the conduct of educational activities. Disadvantaged students can also be evaluated from this perspective, because according to Köse (2019), the quality of life of the school for disadvantaged students and teachers is a central predictor of SE.

5. Suggestions

In this study, inferences were made in light of the data from large-scale applications. Prospective studies can also be planned towards collecting more in-depth data by deploying qualitative data collection methods under a qualitative research model that also includes the opinions of oneself, peers, teachers, and parents. For the education policies that the TES should produce toward increasing SE, it is important to reduce and prevent student bullying in schools with sustainable practices. Exclusively, school administrations should focus on student bullying with the help of counselors, diagnose the problems, and take measures in and out of school according to the types of bullying. In order for students to like their lessons, the curriculum should be arranged according to the interests and needs of children. Educating teachers in a more professional way, helping them make their students love their lessons, and respecting their work may increase their commitment to school.

Limitation of the Research

The data for this research is limited to the TIMSS Turkey 2019 data of 8th grade students.

Disclosure statement and conflict of interest

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors. There is no conflict of interest between the authors. Both authors contributed equally to the article.

6. References

Ahmed, S., Matta, R., Sanyal, N., & Agarwal, S. (2021). Academic optimism of teachers, school engagement and educational aspirations of students from private and government schools. *Journal of Educational Sciences & Psychology, 11*(2),76-85. https://doi.org/10.51865/JESP.2021.2.09

Atilola, O., Abiri, G., & Ola, B. (2022). The impact of behavioral disorders on the level of custodial school-engagement among detained adolescent boys: an observational cohort study. *Journal of Global Health Neurology and Psychiatry*, 33064,1-15 https://doi.org/10.52872/001c.33064

Atoum, A., Al-jarrah, A., & Al Shalalfeh, M. (2019). Parental involvement and school engagement among Jordanian high-primary school students. *Open Science Journal of Psychology, 6*(2), 17-23. http://www.openscienceonline.com/journal/osjp

Aulia, F. (2016). Bullying experience in primary school children. *Indonesian Journal of School Counseling, 1*(1), 28-32. https://doi.org/10.23916/schoulid.v1i1.37.28-32

Ay, Y. (2022). *The effect of mathematics instruction based on flow theory principles enhanced with realistic mathematics education on students 'achievement, retention, motivation and flow state* [Doctoral dissertation]. Middle East Technical University, Ankara.

Biau, G. & Scornet, E., (2016). A random forest guided tour. *An Official Journal of the Spanish Society of Statistics and Operations Research. 25*(2). https://doi.org/10.1007/s11749-016-0481-7

Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics* (Vol. I-II package). CRC Press.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Last, M., & Rice, J. (2003). *Random forests: Finding quasars*. Springer.

Cosma, A., Balazsi, R., & Baban, A. (2015). Bullying in Romanian school children: The role of substance abuse. *Psicologia, Saúde e Doenças, 16*(1), 44-56. http://dx.doi.org/10.15309/15psd160206

Demirci, İ., Usta, F., Yıldız, B., & Demirtaş, A. S. (2020). School burnout and school engagement in adolescents: The mediator and moderator role of facebook addiction. *Egitim ve Bilim, 45*(204),1-19. http://dx.doi.org/10.15390/EB.2020.8377

Demirtas-Zorbaz, S., Zorbaz, O., & Kizildag, S. (2018). School engagement and parent attachment as predictors of perceived competence in adolescents. *Journal of Educational Sciences and Psychology, 8(*1),28-43.

Erentaitė, R., Vosylis, R., Gabrialavičiūtė, I., & Raižienė, S. (2018). How does school experience relate to adolescent identity formation over time? Cross-lagged associations between school engagement, school burnout and identity processing styles. *Journal of Youth and Adolescence, 47*(4), 760-774. https://doi.org/10.1007/s10964-017-0806-1

Fernández-Zabala, A., Goñi, E., Camino, I., & Zulaika, L. M. (2016). Family and school context in school engagement. *European Journal of Education and Psychology, 9*(2), 47-55. https://doi.org/10.1016/j.ejeps.2015.09.001

Ferreira, A. J., & Figueiredo, M. A. (2012). Boosting algorithms: A review of methods, theory, and applications. *Ensemble Machine Learning*, 35-85 https://doi.org/10.1007/978-1-4419-9326-7_2

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Geneur, R., Poggi, J.M., Tuleao Malot, C., Villa-Vialaneix, N., (2017). Random Forest for big data. *Big Data Research. 9*, 28-46. https://doi.org/10.1016/j.bdr.2017.07.003

Glantz, S. A., Slinker, B. K., & Neilands, T. B. (1990). *Primer of applied regression and analysis of variance.* Mc Graw Hill. Inc.

Gr¬nmo, L. S., Kjıırnsli, M., & Lie, S. (2004). *Looking for cultural and geographical factors in patterns of responses to TIMSS items.* Papanastasiou C. Ed. Proceedings of the IRC-2004 TIMSS (pp. 99-112). Cyprus, Cyprus University Press.

Gunuc, S. (2014). The relationships between student engagement and their academic achievement. *International Journal on New Trends in Education and Their İmplications, 5*(4), 216-231.

Han, Y., Kang, H. R., Choe, J. W., & Kim, H. (2021). The moderating role of parental support in the relationship between latent profiles of bullying victimization and sense of school belonging: A cross-national comparison. *Children and Youth Services Review*, 122, 105827.

Huffer, F. W., & Park, C. (2020). A simple rule for monitoring the error rate of random forest for classification. *Quantitative Bio-Science, 39*(1), 1-15.

Japelj Pavešić, B., Radović, M., & Brese, F. (2022). *Students' interests, motivation, and self-beliefs*. In Dinaric Perspectives on TIMSS 2019. Springer, Cham. https://doi.org/10.1007/978-3-030-85802-5_4

Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist, 8*, 7–27. https://doi.org/10.1007/BF03340893

Keller, P. S., El-Sheikh, M., Granger, D. A. ve Buckhalt, J. A. (2012). Interactions between salivary cortisol and alphaamylase as predictors of children's cognitive functioning and academic performance. *Physiology & Behavior, 105*(4), 987-995. https://doi.org/10.1016/j.physbeh.2011.11.005

Klieme, E. (2016). *TIMSS 2015 and PISA 2015: How are they related on the country level*. Deutsches Institut für Internationale Pädagogische Forschung.

Köse, E. K. (2019). Examining the factors affecting the school engagement of at-risk students attending disadvantaged schools. *Egitim ve Bilim, 44(*199),321-336. https://doi.org/10.15390/EB.2019.7893

Kukuş, R. & Akto, A. (2022). Investigation the opinions of secondary school students with regard to their school engagement. *The Journal of International Education Science, 9* (31), 38-68. https://doi.org/10.29228/INESJOURNAL.62474

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade?. *Journal of Educational Psychology, 101*(1), 190-206. https://doi.org/10.1037/a0013153

LaRoche, S., Joncas, M., & Foy, P. (2015). *Sample design in TIMSS 2015. In. Methods and procedures: Tımss 2019 technıcal report.* TIMSS & PIRLS Lynch School of Education International Study Center

Leedy, P. D., & Ormrod, J. E. (2005). Practical research : Planning And Design. Pearson Custom.Li, L., Chen, X., & Li, H. (2020). Bullying victimization, school belonging, academic engagement and achievement in adolescents in rural China: A serial mediation model*. Children and Youth Services Review, 113*(104946), 1–8. https://doi.org/10.1016/j.childyouth.2020.104946

Li, W., Gao, W., & Sha, J. (2020). Perceived teacher autonomy support and school engagement of tibetan students in elementary and middle schools: mediating effect of self-efficacy and academic emotions. *Frontiers in Psychology, 11*(50),1-9. https://doi.org/10.3389/fpsyg.2020.00050

Luo, L., Chen, B., Chen, S., & Zhao, Y. (2020). Association between authoritative school climate and school bullying: Moderation by school belonging. *Ярославский педагогический вестник, 1*,(112), 164-170.

Martin, M. O., von Davier, M., & Mullis, I. V. (2020). *Methods and procedures: TIMSS 2019 technical report.* International Association for the Evaluation of Educational Achievement.

Martin, M.O. (1996). Third international mathematics and science study: An overview. In. M.O. Martin & D.L. Kelly (eds.), *Third international mathematics and science study (TIMSS) technical report,* Volume I: Design and development. Boston College.

Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. In. M.O. Martin & D.L. Kelly (eds.), *Third international mathematics and science study (TIMSS) technical report,* Volume I: Design and development. Boston College.

Mehta B. S, Cornell D, Fan X, Gregory A. (2013). Bullying climate and school engagement in ninth-grade students. *Journal of School Health, 83,* 45-52. https://doi.org/10.1111/j.1746-1561.2012.00746.x

Mo, Y. & Singh, K. (2008). Parents' relationships and ınvolvement: Effects on students' school engagement and performance. *RMLE Online, 31*(10),1-11. https://doi.org/10.1080/19404476.2008.11462053

Murray, C. (2009). Parent and teacher relationships as predictors of school engagement and functioning among low-income urban youth. *The Journal of Early Adolescence, 29*(3), 376-404. https://doi.org/10.1177/0272431608322940

Najam, H., & Kashif, S. (2018). Impact of bullying on students' behavioral engagement. *Journal of Health Education Research & Development, 6*(2), 1-8. https://doi.org/10.4172/2380-5439.1000261

O'Farrell, S. L., Morrison, G. M., & Furlong, M. J. (2006). School engagement. *Children's Needs 3,* 45-58.

OECD (2003). *Student engagement at school. A sense of belonging and participation*. OECD Publishing.

OECD (2013). *PISA 2012 Results: Ready to learn: Students' engagement, drive and self-beliefs (Vol. 3).* PISA: OECD Publishing.

Önen, E. (2014). Öğrencinin okula bağlılığı ölçeği: Türk ortaokul ve lise öğrencileri için uyarlama çalışması. *Turkish Psychological Counseling and Guidance Journal, 5* (42),221-234.

Özdamar, K. (2013). *Paket programlar ile istatiksel veri analizi: MINITAB 16-IBM SPSS 21.* Nisan Kitabevi.

Özdemir, N., & Yalçın, M. T. (2019). Examination of the relationships between academic achievement and the variables at the levels of school and students in secondary schools: Two-level path analysis. *Egitim ve Bilim, 44(*200),93-116. http://dx.doi.org/10.15390/EB.2019.8056

Perdue, N. H., Manzeske, D. P., & Estell, D. B. (2009). Early predictors of school engagement: Exploring the role of peer relationships. *Psychology in the Schools, 46*(10), 1084-1097. https://doi.org/10.1002/pits.20446

Probst, P., & Boulesteix, A. L. (2017). To tune or not to tune the number of trees in random forest. *The Journal of Machine Learning Research, 18*(1), 6673-6690. https://jmlr.org/papers/v18/17-269.html

Reschly, A. L., Huebner, E. S., Appleton, J. J., & Antaramian, S. (2008). Engagement as flourishing: The contribution of positive emotions and coping to adolescents' engagement at school and with learning. *Psychology in the Schools, 45,* 419–431. https://doi.org/10.1002/pits.20306

Revuelta, L. R., Lasarte, O. F., Fernández, A. R., & Maya, M. S. (2018). The role of parental socialization styles in school engagement and academic performance. *European Journal of Education and Psychology, 11*(2), 123-139. https://doi.org/10.30552/ejep.v11i2.226

Şahan, E. B. & Özgenel, M. (2021). Pattern of the relationships between second school students' engagement to school and academic achievements and school effectiveness. *The Journal of International Education Science, 27*(8), 86-103. http://dx.doi.org/10.29228/INESJOURNAL.51469

Smith, T. J., Walker, D. A., Chen, H. T., & Hong, Z. R. (2020). Students' sense of school belonging and attitude towards science: A cross-cultural examination. *International Journal of Science and Mathematics Education, 18*(5), 855-867. https://doi.org/10.1007/s10763-019-10002-7

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*(4), 323-348. https://doi.org/10.1037/a0016973

Tabachnick, B. G., & Fidell, L. S. (2015). *Using multivariate statistics* (6th M. Baloğlu, Trans. Ed.). Nobel.

Uçar, S. (2009). A comparative analysis of earth science education in elementary schools in Turkey and in the USA. *Problems of Education in the 21st Century, 11,* 170-182.

Ünlü, A., Evcin, U., Yılmaz, H. B., & Dalkılıc, A. (2013). İstanbul devlet liselerinde suç ve şiddet yaygınlığı: Okul türlerinin karşılaştırılması. *Anatolian Journal of Psychiatry, 14(*2),152-160. https://doi.org/10.5455/apd.34494

Uzun, M. S. (2021). *Ortaokul öğrencilerinin okuldaki psikolojik ihtiyaçları ile okula bağlanma düzeyleri arasındaki ilişkinin incelenmesi* [Yüksek lisans tezi]. Karamanoğlu Mehmetbey Üniversitesi, Karaman.

Veiga, F. H., Oliveira, I. M., & Taveira, M. C. (2014). *Student's engagement in school, academic aspirations, and career exploration of Portuguese adolescents.* Proceedings of inted 2014 conference (pp. 7545-7553), Valencia, Spain. http://hdl.handle.net/10451/10781

Wang, M. T., & Eccles, J. S. (2012). Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development*, *83*(3), 877-895. https://doi.org/10.1111/j.1467-8624.2012.01745.x

Wang, M. T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28,* 12-23. https://doi.org/10.1016/j.learninstruc.2013.04.002

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79-82. https://doi.org/10.3354/cr030079

Willms, J. D. (2003). *Student engagement at school. A sense of belonging and participation*. Organisation for Economic Co-operation and Development

Yang, (2015). *Multi-level associations between bullying victimization, school engagement, and socioemotional learning.* [Doctoral dissertation]. University of Delaware. http://udspace.udel.edu/handle/19716/17463