Education and Science

Vol 47 (2022) No 209 239-259

Original Article

Determination of the Characteristics Predicting Science Achievement through the Classification and Regression Tree (CART) Method: The Case of TIMSS 2015 Turkey

İzzettin Aydoğan [1], Selahattin Gelbal [2]

Abstract

In the current study, it was aimed to determine student, teacher and school characteristics that predict science achievement of eight grade students in Turkey. In the study, the data of TIMMS 2015 were used and the study group was comprised of a total of 6079 students and 220 teachers from 218 different schools. As the data collection tools, the eighth grade science achievement test used in TIMMS 2015 and the scales administered to students and teachers and reflecting student, teacher and school characteristics were used. Since there was a multi-level data structure where students were nested in schools, the created model was analyzed by using the RE-EM algorithm, which enables multi-level data structures to be analyzed through the classification and regression tree (CART) method. The predicted variable of the model was students' science achievement scores and the predictor variables were the seventeen student, teacher and school characteristics expressed in the scales. According to the results obtained, it was determined that five of the seventeen predictor variables predicted the students' science achievement, which are students confident in science, student bullying, teaching limited by student needs, school discipline problems and school emphasis on academic success. It has been observed that students who have students confidence in science, level of bullying they are exposed to, emphasis on academic success in their schools, school discipline problems in their schools are more, and whose teachers stated that they have more teaching limited by student needs, are more successful.

Keywords

Science achievement

Predictor variables

Student, teacher and school characteristics

Classification and regression tree (CART)

RE-EM algorithm

Article Info

Received: 01.14.2020

Accepted: 11.25.2021

Online Published: 01.19.2022

DOI: 10.15390/EB.2022.9368

[1] Hacettepe University, Institute of Educational Sciences, Department of Educational Sciences, Turkey, ayd.izzet@gmail.com

[2] Hacettepe University, Faculty of Education, Department of Educational Sciences, Turkey, gelbal@hacettepe.edu.tr

Introduction

Science is defined as the means of examining and explaining entities and events in a certain field, creating related principles and making predictions through all these efforts. Natural sciences as a branch of science are also concerned with the same purpose and are defined as an effort to systematically examine, research and determine principles related to entities and events in nature. Adaptation of individuals to the environment and living conditions in which they live depends on their learning features such as getting to know the environment they live in well and making inferences by establishing cause-effect relationships between phenomena (Kaptan & Korkmaz, 2001).

Today, which is described as the age of technology, the main goal of education systems should be to equip students with the skills necessary to have access to information rather than transferring information so that they can adapt to the environment they live in. This will only be possible by imparting higher order skills such as comprehension, application, problem solving and process management to students. Science classes come to the fore in terms of imparting these skills to students (Bayrak & Erden, 2007; Kaptan & Korkmaz, 2001). In science classes, the main objective is to enable students to scientifically research and evaluate the environment and conditions in which they live. Individuals who are scientifically literate have features such as using basic moral values, making synthesis, questioning, analyzing natural events, reasoning and being creative (Kaptan & Korkmaz, 2001).

On the other hand, innovations and discoveries in science, which are accepted as the source of all scientific and technological developments, make a significant contribution to the development of countries. Individuals who are taught how to use scientific, dynamic and productive reasoning are labelled as qualified manpower and these people generally introduce innovations to the society in which they live so that the society can meet the needs of its members. Therefore, societies consisting of individuals with the specified characteristics can ensure development and progress in many areas. Again, the role of science education is very important for the effective accomplishment of this process (Hançer, Şensoy, & Yıldırım, 2003).

Given the delineations above, it is a safe conclusion that conducting research in the field of science teaching and enhancing the quality of science education seem to be a necessity (Ayas, 1995; Ayas, Çepni, & Akdeniz, 1993; Bayrak & Erden, 2007). Almost all societies, especially the developed ones, attach great importance to science and technology education and compete with each other to become better (Bayrak & Erden, 2007). In this context, international measurement and evaluation applications have been carried out regularly from past to present in order to examine the quality and impact of science education, to inform curriculum implementers in different countries and to investigate the successes in different applications and to reveal the best (Kılıç, 2002).

TIMSS (Trends in International Mathematics and Science Study) exams are one of the applications performed especially in the fields of science and mathematics in order to evaluate the educational outcomes of countries and to compare them with each other in the world. TIMSS is organized by the Netherlands-based IEA (International Association for the Evaluation of Educational Achievement). TIMSS aims to identify students' multi-faceted skills. It is administered to fourth and eighth grade students. The students who will participate in the study are randomly selected to represent the country, and the study is conducted every four years (Ministry of National Education [MoNE], 2016). About sixty countries have recently participated in TIMSS (Hooper, Mullis, & Martin, 2013). Before each study, the quality, scope and procedure of the measurement tools to be used are determined and organized in meetings attended by field experts and country representatives (MoNE, 2016). In addition to mathematics and science achievement tests, scales designed to measure social and affective characteristics of students, teachers and school environments are used as measurement tools (Hooper et al., 2013; MoNE, 2016).

Given the individual achievements and gains fostered by science literacy and the role it plays in the scientific, technological and social development of countries, the importance of science education and its development is clearly understood. On the other hand, it is seen that international exams allow the necessary evaluations and comparisons to be made both within the country and between countries so that the scope and quality of science education can be improved. In this regard, the results of the current study are thought to contribute to the development of science education as well as to the evaluation and comparison of science education. In the literature, there are studies conducted on the basis of the research problems and data similar to the ones focused on in the current study using the data from PISA (Programme for International Student Assessment) and TIMSS carried out in different years (Anıl, 2009; Atar & Atar, 2012; Beese & Liang, 2010; Chen, Lin, Wang, Lin, & Kao, 2012; Forbes, Neumann, & Schiepe-Tiska, 2020; Gee & Wong, 2012; Kaya & Doğan, 2017; Mohtar, Halim, Samsudin, & Ismail, 2019; Wiberg & Rolfsman, 2019). In these studies, it has been attempted to determine the factors that predict the science achievement of students and the characteristics related to science achievement using analysis techniques such as hierarchical linear model (HLM), structural equation model (SEM), multiple regression, logistic regression, ANOVA and *t* test. One of the contributions of the current study will be the use of the classification and regression tree (CART) method to determine the predictor variables. It is possible to list the advantages of the CART method as presenting the determined predictor elements with an easily understandable visual output, not requiring a parametric data structure unlike other analysis techniques and interpreting the outputs by classifying them with cut-off scores. There are also various studies (Alivernini & Manganelli, 2012; Depren, 2018; Gomes & Jelihovschi, 2019) conducted using the CART method in order to determine the features that predict science achievement. It is thought that the current study, different from the studies mentioned above, will contribute to the literature by analyzing the multi-level data structure, interpreting the outputs through cut-off scores, and with the types of variables included in the model.

In this connection, the current study aims to determine the characteristics of students, teachers and school environments that predict the science achievement of eighth grade students. To this end, an answer to the following research question will be sought; "What are the variables that predict the science achievement of eight grade students in Turkey and the characteristics belonging to these variables?" The data structure analysed depending on the problem statement of the study, the scope of the statistical model constructed to find a solution and the characteristics possessed by the statistical technique outlined above and whose details will be given in the method section are believed to make the current study significant and meaningful.

Method

Research Type

The current study aiming to determine the variables that predict students' science achievement and the characteristics belonging to these variables is a correlational study. In correlational studies, if there is a significant correlation between two variables, one of the variables can be predicted from the other (Fraenkel & Wallen, 2006). If the number of predictor variables in the research is two or more, the research design is expressed as a multi-factor predictive correlational design (Büyüköztürk, Çakmak, Akgün, Karadeniz, & Demirel, 2014).

Study Group

The data of the study were obtained from the responses to TIMSS 2015 and in the selection of the sample, the two-stage stratified sampling method was used. In the first stage, schools were selected and in the second stage, one or two classes of the schools selected in the first stage were randomly determined (LaRoche, Jonkas, & Foy, 2016).

In the study, the data of the eight graders making up the Turkish sample in TIMSS 2015 and the data belonging to the teachers of these students were used. In this context, the study was conducted on 6079 students (48% are females and 52% are males) and 220 teachers (47% are females and 53% are males) from 218 different schools. The total number of eighth grade students in Turkey is 1,187,893 (48.3% are females; 51.7% are males) (MoNE, 2016).

Data Collection Tools

The data collection tools of the study are the science achievement test used in TIMSS 2015 and the scales administered to the students and teachers. The predicted variable of the developed model is the science scores taken by the students from the achievement test and the predictor variables are the total scores obtained from the scales reflecting the characteristics of students, teachers and schools (TIMSS & PIRLS, 2019).

Student characteristics are determined through students' responses to the relevant scale items, teacher characteristics are determined through teachers' responses to relevant scale items and school characteristics are determined through the responses given by both teachers and students to the relevant scale items. Student characteristics included in TIMSS 2015 and included in the current study are expressed in terms of affective domain (students like learning science, students confident in science, students value science), students' views on engaging teaching in science lessons, students' sense of school belonging and student bullying. School and teacher characteristics; on the other hand, are explained as general school resource shortages, resource shortages for science lessons, school discipline problems, problems with school conditions and resources, school emphasis on academic success, teacher job satisfaction, challenges facing teachers, safe and orderly school, teaching limited by student needs, teachers emphasize science investigation and teachers confident in science (Hooper et al., 2013).

Each item in the scales was scored between 1 and 4 points (minimum 1, maximum 4). Therefore, the lowest total score that can be obtained at each scale level is the same as the number of the items in the scale. The scale of resource shortages for science lessons consists of 5 items; each of the scales of students' sense of school belonging, problems with school conditions and resources, teacher job satisfaction and teaching limited by student needs consists of 7 items; each of the scales of students confident in science, challenges facing teachers and safe and orderly school consists of 8 items; each of the scales of students like learning science, students value science, student bullying and general school resource shortages consists of 9 items; each of the scales of students' views on engaging teaching in science lessons and teachers confident in science consists of 10 items; the scale of school discipline problems consists of 11 items; the scale of teachers emphasize science investigation consists of 15 items and the scale of school emphasis on academic success consists of 17 items. On the other hand, only the scale of teaching limited by student needs is in the form of 3-point Likert scale while all the other scales are in the form of 4-point Likert scale. Thus, the highest score to be taken from the scale of resource shortages for science lessons is 20, from the scale of teaching limited by student needs is 21, from the scales of students' sense of school belonging, problems with school conditions and resources and teacher job satisfaction is 28, from the scales of students confident in science, challenges facing teachers and safe and orderly school is 32, from the scales of students like learning science, students value science, student bullying and general school resource shortages is 36, from the scales of students' views on engaging teaching in science lessons and teachers confident in science is 40, from the scale of school discipline problems is 44, from the scale of teachers emphasize science investigation is 60 and from the scale of school emphasis on academic success is 68. Some of the items in some of the scales used in the current study are reverse worded. Responses to these items were evaluated by means of re-coding according to the general structure of the scale items. After the adjustments, the interpretation of the total score obtained from each scale will be as follows: If the scale scores increase, it means that the attribute addressed in the scale is also increasing (Martin et al., 2016).

For each country data included in TIMSS 2015, the reliability of the measurements obtained from the scales was investigated by using the Cronbach alpha technique and the factor structure formed by the items in the scales was investigated by using the principal components analysis method, and the values reached in the Turkish sample, which is the subject of the current study, are given in Table 1 (Martin et al., 2016).

Table 1. Reliabilities of the Scales and Factor Loading Values

Scales	Cronbach alpha	Factor loading	
		Min.	Max.
Students Like Learning Science	.88	.54	.85
Students Confident in Science	.84	.65	.75
Students Value Science	.90	.62	.82
Students' Views on Engaging Teaching in Science Lessons	.91	.58	.83
Students' Sense of School Belonging	.78	.52	.74
Student Bullying	.81	.54	.75
General School Resource Shortages	.88	.53	.71
Resource Shortages for Science Lessons	.88	.43	.80
School Discipline Problems	.95	.71	.88
Problems with School Conditions and Resources	.88	.59	.85
School Emphasis on Academic Success	.89	.46	.78
Teacher Job Satisfaction	.88	.54	.88
Challenges Facing Teachers	.72	.33	.81
Safe and Orderly School	.88	.69	.80
Teaching Limited by Student Needs	.71	.55	.74
Teachers Emphasize Science Investigation	.85	.47	.87
Teachers Confident in Science	.88	.57	.82

The reliability of the measurement values is acceptable if the Cronbach alpha value is greater than .70 (George & Mallery, 2003). The indicator of the fact that the items in the scales represent the factor structure constructed by the scale is factor loadings higher than .30 (Stevens, 2002). When the values in Table 1 are examined, it is seen that the measurement values for each scale meet the reliability criterion and the items in each scale represent the relevant factor structure.

Data Analysis

The predicted variable of the model created for the analysis of the research data is the scores taken by the students from the science achievement test. The predictor variables on the other hand are the scores taken from the seventeen scales; students like learning science, students confident in science, students value science, students' views on engaging teaching in science lessons, students' sense of school belonging, student bullying, general school resource shortages, resource shortages for science lessons, school discipline problems, problems with school conditions and resources, school emphasis on academic success, teacher job satisfaction, challenges facing teachers, safe and orderly school, teaching limited by student needs, teachers emphasize science investigation and teachers confident in science.

Science achievement scores were obtained by averaging five different plausible values calculated in TIMSS 2015. It is emphasized that the use of any of the possible values or their average in secondary analyses may not produce unbiased estimates or may produce false standard errors (OECD, 2009; Rutkowski, Gonzalez, Jonkas, & Von Davier, 2010). However, Benton (2019), who examined the PISA 2015 data of the United Kingdom (UK), stated that although the distributions of the abilities of the English and Northern Irish students evaluated in the same group were very close, their possible value distributions were different and this situation was not consistent with the emphasized view. On the other hand, in their studies, Aparicio, Cordero, and Ortiz (2021); Laukaityte and Wiberg (2017) and Marchant (2015) stated that although the value created by averaging the possible values produces different standard error and variance measures compared to the results obtained by using all other possible values, it can be used as the output variable as the estimations are very close.

In addition, in order to test whether it is appropriate to use the mean of possible scores, the relationship between the regression factor coefficients obtained by the principal components analysis for five possible values and the mean possible scores was analyzed by using the correlation technique. It was determined that there was a statistically significant and excellent correlation (r=1.00, p< .05) between the regression factor coefficients obtained as a result of the principal components analysis of five possible values and the average scores obtained by taking the averages of five possible values (Ceyhan, 2020).

Since the data were nested within clusters, the analysis process was carried out on the basis of the CART method by using the RE-EM (Random Effect-Expectation Maximization) algorithm (Sela & Simonoff, 2012), which is one of the CART-based algorithms developed for the analysis of multi-level data structures. The school level was taken as the clustering criterion of the data. The analysis of the model created in this framework was made through the *R* program nlme (Pinheiro et al., 2021), REEMtree (Sela, Simonoff, & Jing, 2021), rpart (Therneau, Atkinson, & Ripley, 2019), rpart.plot (Milborrow, 2021) packages.

Before the CART analysis, at the stage of preparing the data for analysis, firstly, the same scale items were separated and the items with reverse response options were determined, and in order to make the scale scores more understandable, in some scales, the direction of scoring was changed by using the reverse scoring method (1-4, 2-3, 3-2, 4-1) (Tezbaşaran, 2008). Then, outlier and missing data analyses were made and these values were determined and the necessary arrangements were made. The missing data values reached at each variable level were lower than 5% (between .5% and 2.5% values), and the data completion procedure was preferred (Schafer, 1999). Some of the missing data distributions were observed at random and some of them were observed at non-random (systematic) level, and the missing data were completed with the closest mean value assignment and value assignment with regression equation methods (Karaatlı, 2014). Then, the predictor variables of the CART model were obtained by calculating the total scores for each scale level.

Classification and Regression Tree (CART)

One of the decision trees methods, CART provides the opportunity to solve two types of problems with a single technique by bringing integrity to the traditional decision tree methods used for solving classification or regression type problems (Trendowicz & Jeffery, 2014). Although the CART method does not require any assumptions such as normality, linearity, homogeneity, etc., which are among the assumptions of parametric regression techniques, it is considered to be an alternative to the multivariate regression analysis when the predicted variable is continuous and to the logistic regression analysis, when the predicted variable is categorical (binary) (Kayri & Boysan, 2008). CART is given different names according to the structure of the predicted variable; Regression Tree (RT) when the predicted variable is continuous, and Classification Tree (CT) when it is categorical (Chang & Wang, 2006). The decision tree structure formed by the CART method is given in Figure 1.

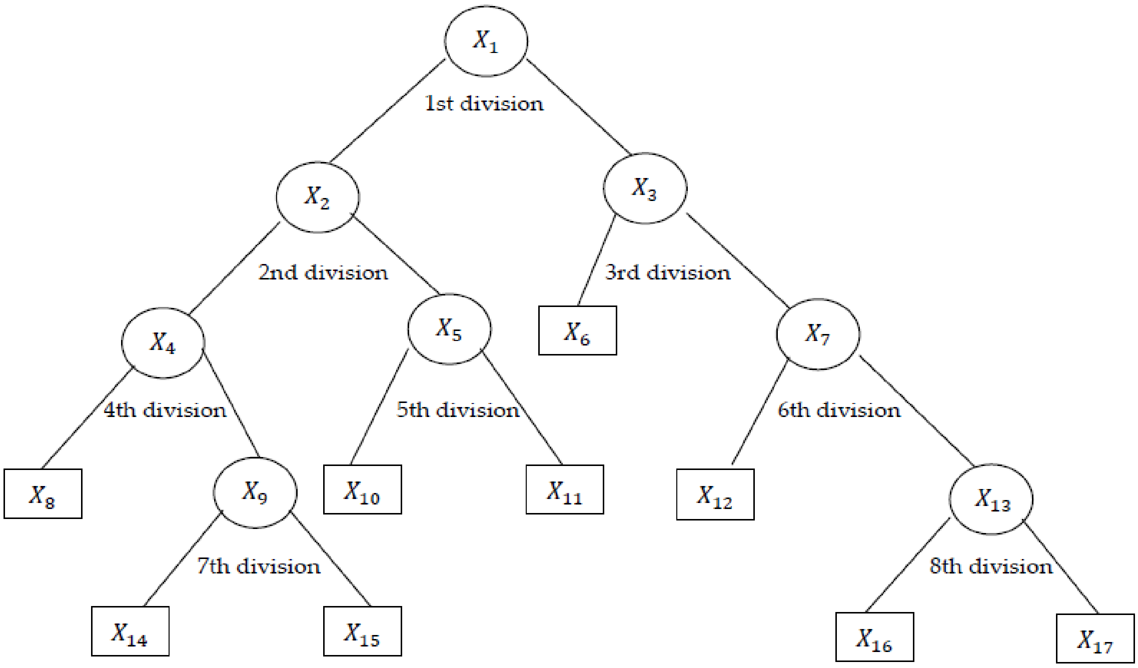

Figure 1. CART Method Tree Structure (Breiman, Friedman, Olshen, & Stone, 1984, p. 21)

The CART method is based on a modelling similar to tree branching in which predictor variables seen to be related to the predicted variable are ordered from top to bottom with the strongest one at the top. Variables that are argued to exhibit significant and strong differentiation between observations are expressed with structures defined as nodes. It is ensured that the observations are divided into two groups with a cut-off point according to the characteristics of the variables at the nodal points, homogeneous within themselves and heterogeneous between groups. These divisions are expressed as branches in the tree structure. The assignment of observations to groups is based on calculations made on the basis of the probability of their being included in that group. At the nodes in the lower steps, these processes are repeated and the division ends when it comes to the variable in which the significant and strong differentiation between the observations ends; thus, the tree structure is completed (Breiman et al., 1984).

Not requiring assumptions required by parametric regression methods such as normality, homogeneity, linearity, etc., showing successful performance in cases where parametric methods cannot produce solutions, having algorithms that take into account linear or non-linear interactions that may occur between variables, not requiring operations during the preparation of the data for analysis as in parametric methods are among the advantages of the CART method (Yamauchi, Ono, Baba, & Ikegami, 2001). Completion of the missing data in data groups on the basis of imitating similar data properties is the means used by the method to handle missing data (Speybroeck, 2012). In addition, it is a method that is resistant to outliers (extreme values) (Timofeev, 2004). Expressing the model with a diagram output (shape) is seen as another advantage as it makes it easier to visualize and interpret (Breiman et al., 1984). There are some disadvantages of the method including profound changes that can be encountered as subtracting or adding few observations from or to the data set and that can change the learning structure and production of divisions and nodes more than necessary in cases where good control cannot be established for the sake of forming classifications in the data set (Timofeev, 2004).

In the CART methodology, the tree formation and development process takes place in three stages (Liaw & Wiener, 2002; Timofeev, 2004):

1. Reaching the strongest tree structure with adjustments,
2. Determination of the size of the tree,
3. Classification of new data using the developed tree.

In the first stage, the best divisions need to occur to create the strong tree structure. The division algorithms differ for classification trees (CT) and regression trees (RT). Divisions in classification trees aim to ensure the maximum homogeneity of the groups that will be located at the nodes (Timofeev, 2004). Common algorithms used as impurity functions aiming to transform the nodes where the observations will be located into a homogeneous structure in classification trees are misclassification error, gini, twoing and entropy indexes (Moisen, 2008). In regression trees, division occurs through algorithms that minimize the expected total variances for the two obtained nodes (Timofeev, 2004). Common algorithms used in regression trees to minimize the variance of the observed values at the nodes are the least squares and the least absolute deviations indexes (Moisen, 2008).

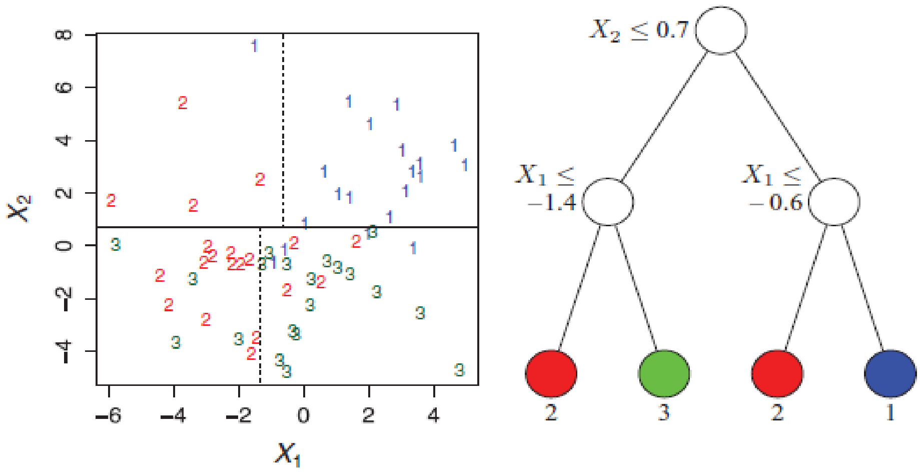

Figure 2. Division in the CART Method and Classification of Observations (Loh, 2011, p. 14)

The details of how the nodes are divided in the CART method and how the observations are located on the nodes are explained with the help of the visual presented in Figure 2. When the index value calculated for an x observation located in the X_2 node is ≤ .7, it is assigned to the left X_1 node and when it is > .7, then it is assigned to the right X_1 node and thus, the division takes place. If the x observation is located in the left X_1 node and the calculated index value is ≤ -1.4, then it is assigned to the node with number 2 and if it is > -1.4, then it is located in the node with number 3. If the x observation is located in the right X_1 node and the calculated index value is ≤ -.6, then it is located in the node with number 2 and if it is > -.6, then it is located in the node with number 1 (Loh, 2011).

The resulting tree structure can be quite complex or consist of many nodes. However, problems such as over-learning or producing low-level predictions can be encountered. In the second stage, it is aimed to provide the best tree size. There are two commonly used methods to obtain the appropriate tree size. The first of these methods is to determine a threshold value, which expresses the homogeneity level of the nodes and under which no division can be made. The other method is to prune an overgrowing tree up to the minimum number of nodes to ensure optimal size (Moisen, 2008). In order to reach the optimal size, the minimum number of observations in each node is taken into account or cross-validation can be applied (Timofeev, 2004).

Another method used in the improvement of the tree is the bagging method. Despite pruning, observations may still have a high variance or deviation tendency. In the bagging method, new data sets with the same distribution are obtained from the data set used as the training set and multiple tree models are created by considering the basic model for each data set. Multiple models are combined by taking the schemas of the tree models for classification and the averages for regression. Thus, the obtained values are processed and excessive deviation or variance values are controlled and minimized (Moisen, 2008).

In the third stage, new data are classified and assigned to the appropriate nodes. It is aimed to locate the new data in the existing classes/nodes with the predictive power of the tree structure created with this process (Liaw & Wiener, 2002; Timofeev, 2004).

RE-EM Algorithm for the Multi-Level CART Model

Although it seems to be appropriate due to its non-parametric nature, its tendency to select variables with higher division potential does not make the standard CART method suitable for the analysis of multilevel data groups. In the standard CART method, the selection of the variables at different levels is based on the division potentials, regardless of at which level they are. For example, a variable with n observations will have the n-1 division potential at the first level, whereas if the same variable is nested in k clusters, it will have the k-1 division potential. In this case, this variable's being at the first level will increase its probability of being selected (Enfield, 2015).

One of the tree-based algorithms developed for multi-level CART models and proposed by Sela and Simanoff (2012) is the RE-EM algorithm which allows analyzing clustered, longitudinal data structures with the regression tree method and which repeats the EM (Expectation Maximization) (Laid & Ware, 1982) algorithm. This procedure works on the basis of the principle of estimating random effects from the iterations of the tree structure during its formation (as cited in Loh, 2014). Since both random effects and fixed effects are not known, it is iterate between estimated the regression tree, assuming that estimates of the random effects are correct, and estimating the random effects, assuming that the regression tree for the fixed effects is correct (Fu & Simonoff, 2015). The tree structure processed in the RE-EM algorithm is based on the CART methodology (Fu & Simonoff, 2015; Mancini & Sacco, 2020).

Results

The findings obtained from the analysis of the data collected to find an answer to the problem statement of the study "What are the variables that predict the science achievement of eight grade students in Turkey and the characteristics belonging to these variables?" within the context of the variables reflecting student, teacher and school characteristics and the findings obtained from the analysis of the CART model constructed are presented below.

Since the CART method is methodologically built on the learning-based logic, the analyses were conducted by dividing the data into two as the training set (*n*=4246) constituted by 152 (70%) school data and the test set (*n*=1833) constituted by 66 (30%) school data according to the school level taken as the clustering criterion (Ahmad, Reynolds, & Rezgui, 2018; Ahmed & Elaraby, 2014; Pahmi, Saepudin, Maesarah, Solehudin, & Wulandari, 2018). The tree was constructed on the grounds that the division would continue as long as the rate of variability explained (complexity value (cp): complexity parameter) was at least .001 and the number of observations in the node was at least 20. In order to prevent over-learning or over-fitting problems, pruning process based on the 10-fold cross validation method was applied. The final tree structure was reached at the value where the cross-validation error was minimum and the complexity parameter was maximum (Sela & Simonoff, 2012).

The predictiveness of the tree model obtained by training with the training data set (*n*=4246) on the new data was determined with the RMSE value. The RMSE value for the predictiveness of the model on the whole data set (*n*=6079) was found to be 64.91 and the RMSE value for its predictiveness on the test data set (*n*=1833) was found to be 65.75. According to the determined variance and error values, the most suitable tree model can be obtained with a structure consisting of 7 divisions and 8 nodes.

The descriptive features of the scale scores expressing the predictor variables of the model created in the current study are presented in Table 2, and in the lowest score column, the minimum total score values from each scale are presented; in the highest score column, the maximum total score values from each scale are presented; in the mean scores column, the total mean scores of the scales are presented and in the standard deviation column, the standard deviation values of the scale scores are presented.

Table 2. Descriptive Features of the Scale Scores

Variable	The lowest score	The highest score	Mean	Standard deviation
Students Like Learning Science	9	36	30.29	5.61
Students Confident in Science	8	32	24.07	5.36
Students Value Science	9	36	29.28	6.26
Students' Views on Engaging Teaching in Science Lessons	10	40	34.80	6.20
Students' Sense of School Belonging	7	28	24.20	3.71
Student Bullying	9	36	12.99	4.87
General School Resource Shortages	9	36	27.94	5.49
Resource Shortages for Science Lessons	5	20	14.25	3.16
School Discipline Problems	11	44	25.26	9.07
Problems with School Conditions and Resources	7	28	17.92	5.42
School Emphasis on Academic Success	17	65	43.67	6.92
Teacher Job Satisfaction	7	28	22.80	4.21
Challenges Facing Teachers	8	30	17.73	4.43
Safe and Orderly School	8	32	23.92	4.90
Teaching Limited by Student Needs	7	20	13.90	2.38
Teachers Emphasize Science Investigation	15	60	41.95	6.99
Teachers Confident in Science	10	40	30.69	5

With the CART model, it was investigated which of the seventeen characteristics of students, teachers and schools predicted students' science achievement. As a result of the analysis, it was determined that five of these seventeen characteristics significantly predicted the predicted variable. In this sense, it was determined that the variables that predicted the science achievement of the eighth grade students were students confident in science, student bullying, teaching limited by student needs, school discipline problems and school emphasis on academic success. The tree diagram showing the analysis results of the variables of student, teacher and school characteristics that predict science achievement investigated by the CART technique is presented in Figure 3.

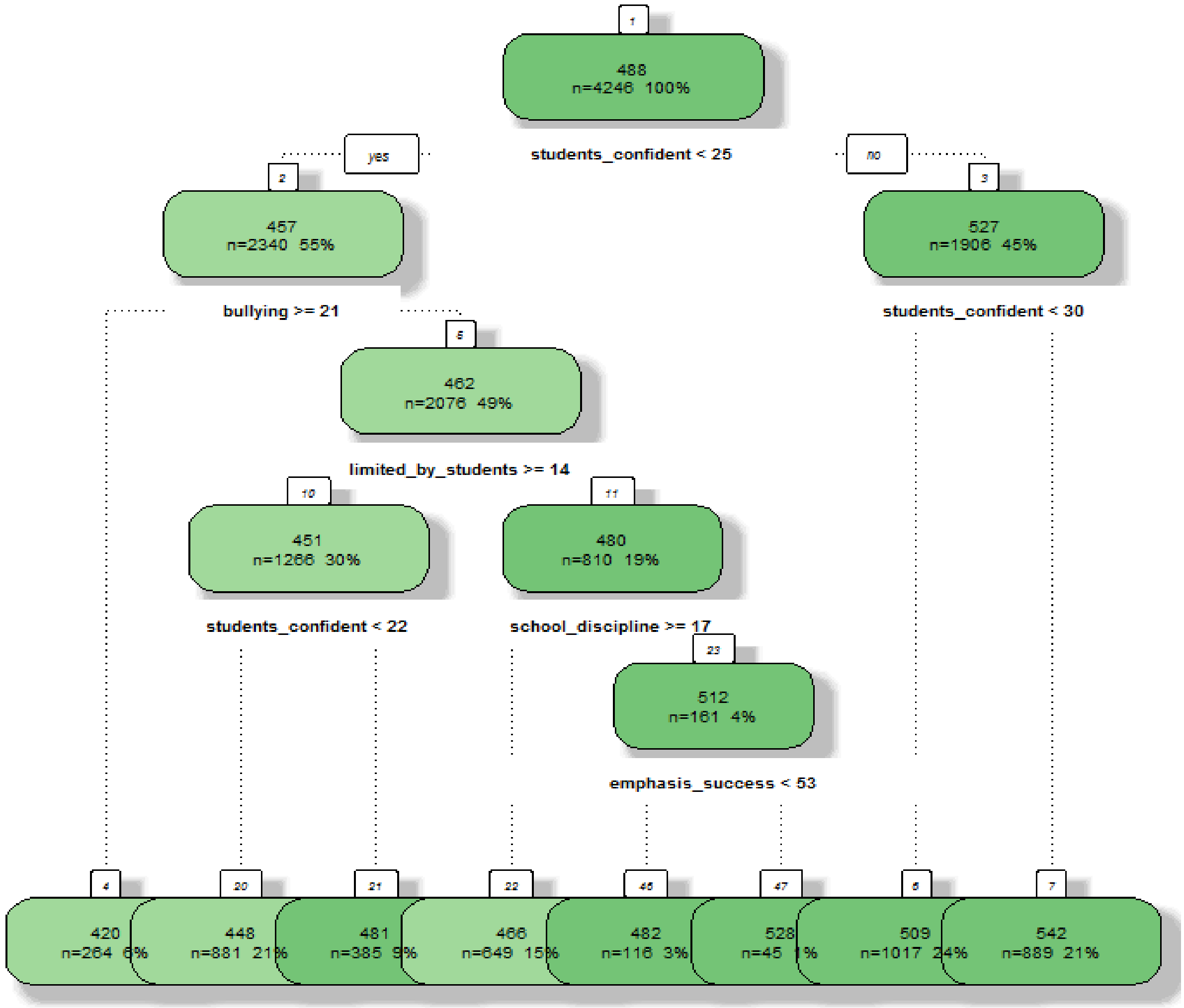

Figure 3. Modelled Regression Tree Diagram

When the results of the CART analysis are evaluated, it is seen that the most important characteristic that predicts the students' science achievement is students confident in science. According to this characteristic, the students were divided into two groups on the basis of the 25 cut-off point and 45% (*n*=1906) and 55% (*n*=2340) slices of the number of observations. The mean science score of the group with a higher score from the scale of students confident in science (≥25) was found to be 527 while that of the group with a lower score (<25) was found to be 457. This indicates that the students with a higher score from the scale of students confident in science have a higher mean science score.

The students with high scores from the scale of students confident in science (≥25) again differed in terms of their self-efficacy belief. The students in this group were divided into two groups on the basis of the 30 cut-off point and 24% (*n*=1017) and 21% (*n*=889) slices of the number of observations. The mean science score of the students in the upper group (≥30) was found to be 542 while the mean science score of the students in the lower group (<30) was found to be 509. It was observed that the students with high scores from the scale of students confident in science were located in the group with the highest mean science score.

The students with low scores from the scale of students confident in science (<25) differed in terms of the variable of student bullying in a sub-node. According to the scale of student bullying, the students were divided into two groups on the basis of the 21 cut-off point and 6% (*n*=264) and 49% (*n*=2076) slices of the number of observations. The mean science score of the students in the upper group

(≥21) who stated that they were exposed to more bullying was found to be 462 while the mean science score of the students in the lower group (<21) who stated that they were exposed to less bullying was found to be 420. It was observed that the students who stated that they were exposed to less bullying were in the group with the lowest mean science score. In this case, it is understood that the mean science score of the students who are exposed to more bullying is higher.

The students taking high scores from the scale of student bullying (≥21) were also observed to differ in terms of the variable of teaching limited by student needs. According to the scale of teaching limited by student needs, the students were divide into two groups on the basis of the 14 cut-off point and 19% (n=810) and 30% (n=1266) slices of the number of observations. The mean science score of the students in the lower group (<14); that is, the students with whom their teachers experienced fewer problems, was found to be 451 while the mean science score of the students in the upper group (≥14), that is, the students with whom their teachers experienced more problems, was found to be 480. These values show that the mean science scores of the students whose teachers stated that they have more problems with them are higher.

The students taking low scores from the scale of teaching limited by student needs (<14) differed in terms of the variable of students confident in science. The students in this group were divided into two groups on the basis of the 22 cut-off point and 9% (n=385) and 21% (n=88) slices of the number of observations. The mean science score of the students in the upper group (≥22) was found to be 481 while the mean science score of the students in the lower group (<22) was found to be 448. Thus, it can be argued that the mean science scores of the students with higher scores from the scale of students confident in science in this group are higher.

The students having higher scores from the scale of teaching limited with student needs (≥14) differed in terms of the variable of school discipline problems. The students in this group were divided into two groups on the basis of the 17 cut-off point and 4% (n=161) and 15% (n=649) slices of the number of observations. The mean science score of the students in the upper group (≥17); that is, the students whose teachers stated that they have more discipline problems at school, was found to be 512 while the mean science score of the students in the lower group (<17); that is, the students whose teachers stated that they have fewer discipline problems at school, was found to be 466. According to this result, it can be stated that the mean science scores of the students whose teachers said that they have more discipline problems at school are higher.

The students with higher scores from the scale of school discipline problems (≥17) were also found to differ in terms of the variable of school emphasis on academic success. The students in this group were divided into two groups on the basis of the 53 cut-off point score and 1% (n=45) and 3% (n=116) slices of the number of observations. The mean science score of the students in the upper group (≥53); that is, the students whose teachers stated that they have more emphasis on academic success at school, was found to be 528 while the mean science score of the students in the lower group (<53); that is, the students whose teachers stated that they have fewer emphasis on academic success at school, was found to be 482. In this context, it can be said that the mean science scores of the students thinking that much emphasis is put on academic success at school are higher.

The variables of students confident in science, student bullying, teaching limited by student needs, school discipline problems and school emphasis on academic success, which were determine to significantly predict students' science achievement with the CART model constructed, were found to explain 23.3% ($R^2 = .233$) of the variance in the science scores.

Conclusion, Discussion and Suggestions

Through the model created by using the CART method, it was investigated which of the seventeen student, teacher and school characteristics related to science achievement predicted the students' science achievement. As a result of the analysis, it was determined that five of these seventeen characteristics significantly predicted science achievement while the other twelve characteristics did not significantly predict science achievement.

The variable that came to the fore in the prediction of the students' science achievement was observed to be the variable of students confident in science. The students with higher scores taken from the scale of students confident in science; that is, the students having higher self-efficacy in science, were observed to be more successful in science. When the cut-off points produced by the model while classifying the observations were evaluated, it was found that the first division in the scores taken from the scale of students confident in science was performed with the 25 cut-off point, the second division with the 30 cut-off point and the third division with the 22 cut-off point. Given that the highest score to be taken from the scale is 32 and the mean score is 24.07, it can be argued that the cut-off points are close to or over the mean; accordingly, that having a moderate or higher self-efficacy in science is an important criterion for science achievement. When the research on similar research problems and similar students in the literature is examined, it is seen that similar results regarding the self-efficacy in science have been reported by Acar and Öğretmen (2012); Çalışkan (2008) investigating the PISA 2006 data, by Atar and Atar (2012); Bayraktar (2010) investigating the TIMSS 2007 data and by Batı, Yetişir, and Güneş (2019); Ötken (2019) investigating the PISA 2015 data, they also reported that self-efficacy in science positively affected students' science achievement. However, Ceylan and Berberoğlu (2007) found that there is a negative correlation between self-efficacy belief and science achievement in their study using the TIMSS 2007 data. Performing the analyses with the same data and model but with different statistical programs, Acar and Öğretmen (2012) observed a statistically significant correlation between science achievement and self-efficacy in one, but did not find such a significant correlation in the other. Anagün (2011) conducted a study with the PISA 2006 data and Atar (2014) conducted a study with the TIMSS 2011 data and both of them concluded that self-efficacy belief does not significantly predict science achievement. On the other hand, Hwang, Choi, Lee, Culver, and Hutchison (2016); Komarraju and Nadler (2013); Köseoğlu (2015); Liu, Hsieh, Cho, and Schallert (2006) and Motlagh, Amrai, Yazdani, Abderahim, and Sourie (2011) conducted studies showing that self-efficacy perception positively predicts academic achievement. Based on the conclusion that students with high self-efficacy beliefs are more successful in science lessons, it is thought that it is necessary to give importance to the implementation of classroom practices, methods and activities that will strengthen the sense of success and confidence towards science lessons in schools.

The second variable that predicts students' science achievement is the variable of student bullying. It was determined that the students who stated that they were exposed to bullying more at school were more successful than the students who stated that they were exposed to bullying less. When the cut-off points produced by the model while classifying the observations were evaluated, it was found that the division in the scores taken from the scale of student bullying was performed with the 21 cut-off point. The highest score to be taken from the scale is 36 and the mean score of the scale was found to be 12.99. According to these values, it can be said that students who are successful in science classes are more exposed to peer bullying taking place in different forms such as exclusion, teasing, scolding, violence and threatening. The fact that successful students are exposed to bullying more frequently indicates that bullying behaviours are exhibited by low-achieving students, so students with low academic achievement tend to bully their peers more. Topçu, Erbilgin, and Arıkan (2016), in their study based on the TIMSS 2011 data, emphasized that bullying negatively predicted students' science achievement and students who were bullied were more unsuccessful. In the literature, it is possible to come across studies (Huang, 2020; van der Werf, 2014) that support the findings of Topçu et al. (2016). However, in the study conducted by Özer, Totan, and Atik (2011) on secondary school students, it was revealed that students with low academic achievement show a greater tendency to bully or to be bullied.

Glew, Fan, Katon, Rivara, and Kernic (2005); Konishi, Konishi, Hymel, Zumbo, and Li (2010) and Totura, Green, Karver, and Gesten (2009) stated that students who engage in bullying are academically less successful. Therefore, it is understood that the relationship between bullying which is expressed in terms of being bullied and exhibiting bullying behaviors, and academic achievement may be in different directions. In this context, in order to increase students' science achievement, it is thought that it will be important for school administrators to take preventive measures against bullying, to inform students frequently about the effects and outcomes of bullying and for school counselling services to provide students with necessary support including the educational process.

Another variable that was found to predict the students' science achievement in a statistically significant manner is the variable of teaching limited by student needs. It is understood that the students whose teachers stated that they experience problems with the characteristics and behaviours of these students are more successful. When the cut-off point produced by the model while classifying the observations was evaluated, it was observed that the division in the scores taken from the scale of teaching limited to student needs was performed with the 14 cut-off point. Given that the highest score to be taken from the scale is 21 and the mean score taken from the scale is 13.9, it can be argued that the students whose teachers stated that they experience more problems with these students are more successful in science lessons. This finding does not concur with the literature. Bayraktar (2010) conducted a study by using the TIMSS 2007 data and Üstün, Özdemir, Cansız, and Cansız (2020) conducted a study by using the PISA 2015 data and both of them found that with students' decreasing level of creating problems for their teachers, indifference to classes and lack of preparedness, their science achievement increased. The studies conducted by Amrai, Motlagh, Zalani, and Parhon (2011); Barriga et al. (2002) and Singh, Graville, and Dika (2002) also support the general view in the literature. It is thought that one of the important reasons for obtaining such a result in the current study can be related to the fact that this variable can highly change depending on teacher attitudes and perceptions. As stated by Geving (2007), Gorham and Cristophel (1992) and Jarvis and Seifert (2002) factors such as teachers' behaviours and attitudes towards students, teaching methods and techniques and competency levels can cause students to exhibit behaviours such as low motivation, apathy, shyness, avoidance of academic responsibility and helplessness. And, such behaviours in students can be considered as a problem by their teachers. For this reason, beyond the perceptions of teachers, it is thought that such behaviours of students and their science achievement may be correlated with different psychological factors.

Another variable that was found to significantly predict students' science achievement is the variable of school discipline problems. The students who were stated to have fewer discipline problems by their teachers were found to be more successful. When the cut-off value produced by the model while classifying the observations was evaluated, the division in the scores taken from the scale of school discipline problems was performed with the 17 cut-off point. Given that the highest score to be taken from the scale is 44 and the mean score of the scale is 23.92, it can be argued that students with fewer discipline problems are more successful in science. Abazoğlu, Yıldızhan, and Yıldırım (2014) reached a similar conclusion in their study conducted with the TIMSS 2011 data. In the study of Atar (2014), on the other hand, the school safety variable was not found to be a statistically significant predictor of science achievement. Grover (2015); Kim, Sanders, Makubuya, and Yu (2020); Lassen, Steele, and Sailor (2006) and Whisman and Hammer (2014) concluded in their studies that, parallel to the findings of the current study, students are more successful in schools where the environment is safe and disciplinary problems are rarely experienced. As stated by Osher, Bear, Sprague, and Doyle (2010) and Skiba and Peterson (2003), in order to ensure a safe environment and control of disciplinary problems in schools, working in coordination with the family, acting in such a way as to increase students' commitment to school rather than punishing or excluding them and implementing regular psychological support programs for students will contribute to the improvement of science success.

The last variable that was found to significantly predict the students' science achievement is the variable of school emphasis on academic success. It was determined that students studying in schools where success is considered to be important and extra studies are carried out to support success are more successful than students in schools where such an importance is not attached to academic success. When the cut-off point produced by the model while classifying the observations was evaluated, it was observed that the division in the scores taken from the scale of school emphasis on academic success was performed with the 53 cut-off point. Given that the highest score to be taken from the scale is 68 and the mean score of the scale is 41.95, it can be argued that students in schools where academic success is supported more than the average and where more than average activities are conducted to support academic success are more successful in science classes. In the study conducted by Atar (2014) using the TIMSS 2011 data, similar results were reported. In general, other studies in the literature (Chen & Wong, 2015a, 2015b; Filippello, Buzzai, Costa, Orecchio, & Sorrenti, 2020; Tanaka & Yamauchi, 2000) support the view that academic performance increases as the importance given to success increases. In this connection, it is thought that works such as directing students to tasks and responsibilities in which they can experience success, ensuring that students feel the sense of success, and making reinforcing academic activities more widespread will contribute to the development of students' science achievement.

On the other hand, Anagün (2011) reached a similar result using the PISA 2006 data for the variable of students like learning science (attitude), which was not found to be statistically significant as a predictor of students' science achievement in the current study. Ceylan and Berberoğlu (2007) found a negative correlation between them and concluded that students with stronger attitudes are less successful. Anıl (2009) conducted a study by using the PISA 2006 data, Kaya and Kaya (2019) conducted a study by using the TIMSS 2015 data and Uzun, Gelbal, and Öğretmen (2010) conducted a study by using the TIMSS 1999 data and all of them concluded that attitude positively predicts science achievement. In addition, Acar and Öğretmen (2012); Batı et al. (2019); Bayraktar (2010) and Ötken (2019) reached similar results in their studies. In the current study, it was observed that the value attached to science does not significantly predict science achievement. However, Batı et al. (2019); Ceylan and Berberoğlu (2007); Çalışkan (2008) and Uzun et al. (2010) concluded that the value attached to science significantly predicts science achievement. Similarly, the variable of students' views on engaging teaching in science lessons found to be insignificant in terms of predicting students' science achievement was also found to be insignificant by Uzun et al. (2010); the variable of students' sense of school belonging found to be insignificant in terms of predicting students' science achievement in the current study was also found to be insignificant by Topçu et al. (2016) and the variable of teacher job satisfaction found to be insignificant in terms of predicting students' science achievement in the current study was also found to be insignificant by Abazoğlu and Taşar (2016) and Atar (2014).

When the relevant studies in the literature are examined, it is understood that students who have high self-efficacy beliefs, who study in schools where success is considered to be important, where the environment is safe and where discipline is highly estimated are more successful in science lessons, consistent with the findings of the current study. However, the finding that the students who are in schools where more problems arising from students are experienced are more successful does not concur with the literature. In parallel and in contrast to the conclusion that students who are exposed to bullying are more successful, it is possible to come across different results regarding the effect of bullying on academic success. In addition, it is observed that students' attitudes towards science lesson and the value attached to science, which were found to be insignificant in terms of predicting academic achievement, were mostly found to be significant in the literature.

When the limitations of the CART method used in the current study and the fact that even small changes in the observation or data structure affect the outputs in the divisions and nodes are considered, it becomes clear that in studies that will use this method, the proper check of the accuracy of observations and data structure should be ensured. Furthermore, since it is desired to reach the optimal tree structure in order to provide the best estimation, it will be useful to point out that it is necessary to apply optimization processes such as prevention of over-learning, cross validation, pruning and bagging.

References

Abazoğlu, İ., & Taşar, M. F. (2016). Fen bilgisi öğretmen özelliklerinin öğrenci fen başarısı ile ilişkisi: TIMSS 2011 verilerine göre bir durum analizi. *İlköğretim Online, 15*(3), 922-945.

Abazoğlu, İ., Yıldızhan, Y., & Yıldırım, O. (2014). TIMSS 2011 Türkiye 8. sınıf fen bilimleri sonuçlarının değerlendirilmesi. *Eğitim ve Öğretim Araştırmaları Dergisi, 3*(1), 278-288.

Acar, T., & Öğretmen, T. (2012). Çok düzeyli istatistiksel yöntemler ile 2006 PISA fen bilimleri performansının incelenmesi. *Eğitim ve Bilim, 37*(163), 178-189.

Ahmad, M., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*, *203*, 810-821.

Ahmed, A. B., & Elaraby, I. S. (2014). Data mining: A prediction for student's performance using classification method. *World Journal of Computer Application and Technology*, *2*(2), 43-47.

Alivernini, F., & Manganelli, S. (2012). Science competencies across PISA OECD countries: Comparing exceptionally high and low performers. In *Improving education through accountability and evaluation lessons from around the world 3-5 October 2012.* Center for American Studies, Roma. Retrieved from http://www.umdcipe.org/conferences/EducationEvaluationItaly/ItalyPreliminaryAgendaPage.html

Amrai, K., Motlagh, S. E., Zalani, H. A., & Parhon, H. (2011). The relationship between academic motivation and academic achievement students. *Procedia Social and Behavioral Sciences*, *15*, 399-402.

Anagün, Ş. S. (2011). PISA 2006 sonuçlarına göre öğretme-öğrenme süreci değişkenlerinin öğrencilerin fen okuryazarlıklarına etkisi. *Eğitim ve Bilim*, *36*(162), 84-102.

Anıl, D. (2009). Uluslararası öğrenci başarılarını değerlendirme programı (PISA)'nda Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler. *Eğitim ve Bilim, 34*(152), 87-100.

Aparicio, J., Cordero, J. M., & Ortiz, L. (2021). Efficiency analysis with educational data: How to deal with plausible values from international large-scale assessments. *Mathematics, 9*(1579), 1-16.

Atar, H. Y. (2014). Öğretmen niteliklerinin TIMSS 2011 fen başarısına çok düzeyli etkileri. *Eğitim ve Bilim, 39*(172), 121-137.

Atar, H. Y., & Atar, B. (2012), Examinig the effects of Turkish education reform on student's TIMSS 2007 science achievements. *Educational Sciences: Theory & Practice, 12*(4), 2632-2636.

Ayas, A. (1995). Fen bilimlerinde program geliştirme ve uygulama teknikleri üzerine bir çalışma: Çağdaş yaklaşımın değerlendirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *11*(11), 149-155.

Ayas, A., Çepni, S. & Akdeniz, A. R. (1993). Development of the Turkish secondary science curriculum. *Science Education*, *77*(4), 433-440.

Barriga, A. Q., Doran, J. W., Newell, S. B., Morrison, E. M., Barbetti, V., & Robbins, B. D. (2002). Relationships between problem behaviors and academic achievement in adolescents: The unique role of attention problems. *Journal of Emotional and Behavioral Disorders, 10*(4), 233-240.

Batı, K., Yetişir, M. İ., & Güneş, G. (2019). PISA 2015 sonuçlarına göre duyuşsal özelliklerin öğrencilerin fen performansları üzerine etkisinin ülkeler arası karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 34*(4), 999-1014.

Bayrak, B., & Erden, A. M. (2007). Fen bilgisi öğretimi programının değerlendirilmesi. *Kastamonu Eğitim Dergisi*, *15*(1), 137-154.

Bayraktar, Ş. (2010). Uluslararası fen ve matematik çalışması (TIMSS 2007) sonuçlarına göre Türkiye'de fen eğitiminin durumu: Fen başarısını etkileyen faktörler. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, *30*, 249-270.

Beese, J., & Liang, X. (2010). Do resources matter? PISA science achievement comparisons between students in the United States, Canada and Finland. *Improving Schools, 13*(3), 266-279.

Benton, T. (2019). The effect of using principal components to create plausible values. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative psychology: The 83rd annual meeting of the psychometric society* (pp. 296-303). New York: Springer.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2014). *Bilimsel araştırma yöntemleri* (16th ed.). Ankara: Pegem.

Ceyhan, G. (2020). *Sınıflandırmada kullanılan veri madenciliği yöntemlerinin performanslarının veri seti özelliklerine göre karşılaştırılması* (Unpublished doctoral dissertation). Gazi University, Ankara.

Ceylan, E., & Berberoğlu, G. (2007). Öğrencilerin fen başarısını açıklayan etmenler: Bir modelleme çalışması. *Eğitim ve Bilim, 32*(144), 36-48.

Chang, L. Y., & Wang, H. W. (2006). Analysis of traffic injury: An application of non-parametric classification tree techniques. *Accident Analysis Prevention, 38*(5), 1019-1027.

Chen, S.-F., Lin, C.-Y., Wang, J.-R., Lin, S.-W., & Kao, H.-L. (2012). A cross-grade comparison to examine the context effect on the relationships among family resources, school climate, learning participation, science attitude, and science achievement based on TIMSS 2003 in Taiwan. *International Journal of Science Education, 34*(14), 2089-2106.

Chen, W. W., & Wong, Y. L. (2015a). Chinese mindset: Theories of intelligence, goal orientation and academic achievement in Hong Kong students. *Educational Psychology, 35*(6), 714-725.

Chen, W. W., & Wong, Y. L. (2015b). The relationship between goal orientation and academic achievement in Hong Kong: The role of context. *Asia-Pacific Educational Researcher, 24*(1), 169-176.

Çalışkan, M. (2008). *The impact of school and student related factors on scientific literacy skills in the programme for international student assessment-PISA 2006* (Doctoral dissertation). Middle East Technical University, Ankara.

Depren, S. K. (2018). Prediction of students' science achievement: An application of multivariate adaptive regression splines and regression trees. *Journal of Baltic Science Education*, *17*(5), 887-903.

Enfield, D. P. M. (2015). *Efficiently exploring multilevel data with recursive partitioning* (Doctoral dissertation). University of Virginia, Virginia.

Filippello, P., Buzzai, C., Costa, S., Orecchio, S., & Sorrenti, L. (2020). Teaching style and academic achievement: The mediating role of learned helplessness and mastery orientation. *Psychology in the Schools, 57*(1), 5-16.

Forbes, C. T., Neumann, K., & Schiepe-Tiska, A. (2020). Patterns of inquiry-based science instruction and student science achievement in PISA 2015. *International Journal of Science Education, 42*(5), 783-806.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.

Fu, W., & Simonoff, J. S. (2015). Unbiased regression trees for longitudinal and clustered data. *Computational Statistics and Data Analysis*, *88*, 53-74.

Gee, K. A., & Wong, K. K. (2012). A cross national examination of inquiry and its relationship to student performance in science: Evidence from the Program for International Student Assessment (PISA) 2006. *International Journal of Educational Research, 53*, 303-318.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Geving, A. M. (2007). Identifying the types of student and teacher behaviours associated with teacher stress. *Teaching and Teacher Education*, *23*, 624-640.

Glew, G. M., Fan, M. Y., Katon, W., Rivara, F. P., & Kernic, M. A. (2005). Bullying, psychosocial adjustment and academic performance in elementary school. *Archives of Pediatrics & Adolescent Medicine*, *159*(11), 1026-1031.

Gomes, C. M. A., & Jelihovschi, E. (2019). Presenting the regression tree method and its application in a large-scale educational dataset. *International Journal of Research & Method in Education, 43*(2), 201-221.

Gorham, J., & Christophel, D. M. (1992). Students' perceptions of teacher behaviors as motivating and de-motivating factors in college classes. *Communication Quarterly*, *40*(3), 239-252.

Grover, A. (2015). *Student perception of school safety and how it affects their academic achievement* (Doctoral dissertation). Northwest Nazarene University, Nampa.

Hançer, A. H., Şensoy, Ö., & Yıldırım, H. İ. (2003). İlköğretimde çağdaş fen bilgisi öğretiminin önemi ve nasıl olması gerektiği üzerine bir değerlendirme. *Pamukkale Eğitim Fakültesi Dergisi*, *1*(13), 80-88.

Hooper, M., Mullis, I. V. S., & Martin, M. O. (2013). TIMSS 2015 contex questionnaire framework. I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 assessment frameworks* (pp. 61-82). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf

Huang, L. (2020). Exploring the relationship between school bullying and academic performance: The mediating role of students' sense of belonging at school. *Educational Studies*. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/03055698.2020.1749032

Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D., & Hutchison, B. (2016). The relationship between self-efficacy and academic achievement: A 5-year panel analysis. *Asia-Pacific Education Researcher*, *25*(1), 89-98.

Jarvis, S., & Seifert, T. (2002). Work avoidance as a manifestation of hostility, helplessness or boredom. *Alberta Journal of Educational Research, 48*(2), 174-187.

Kaptan, F., & Korkmaz, H. (2001). *İlköğretimde fen bilgisi öğretimi. İlköğretimde etkili öğrenme ve öğretmen el kitabı modül 7*. Ankara: MEB Yayınları.

Karaatlı, M. (2014). Verilerin düzenlenmesi ve gösterimi. In Ş. Kalaycı (Ed.), *SPSS uygulamalı çok değişkenli istatistik teknikleri* (6th ed., pp. 3-47). Ankara: Asil Yayın Dağıtım.

Kaya, V. H., & Doğan, A. (2017). Determination and comparison of Turkish student characteristics affecting science literacy in Turkey according to PISA 2012. *Research Journal of Business and Management (RJBM)*, *4*(1), 34-51.

Kaya, V. H., & Kaya, E. (2019). Öğrencilerin ortaokul öğrenimleri sürecinde okula, fen bilimleri öğretmenine ve dersine yönelik algıları ile fen başarısı arasındaki ilişki. *Ihlara Eğitim Araştırmaları Dergisi*, *4*(1), 18-36.

Kayri, M., & Boysan, M. (2008). Bilişsel yatkınlık düzeyleri ilişkisinin sınıflandırma ve regresyon ağacı analizi ile incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *34*, 168-177.

Kılıç, B. G. (2002). Dünyada ve Türkiye'de fen öğretimi. In *V. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi 16-18 Eylül 2002*. Ankara: ODTÜ Kültür ve Kongre Merkezi. Retrieved from http://old.fedu.metu.edu.tr/ufbmek-5/b_kitabi/PDF/Fen/Bildiri/t063DA.pdf

Kim, Y. K., Sanders, J. E., Makubuya, T., & Yu, M. (2020). Risk factors of academic performance: Experiences of school violence, school safety concerns and depression by gender. *Child & Youth Care Forum*, *49*(5), 725-742.

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals and effort regulation matter?. *Learning and Individual Differences, 25*, 67-72.

Konishi, C., Hymel, S., Zumbo, B. D., & Li, Z. (2010). Do school bullying and student-teacher relationships matter for academic achievement? A multilevel analysis. *Canadian Journal of School Psychology*, *25*(1), 19-39.

Köseoğlu, Y. (2015). Self-efficacy and academic achievement - A case from Turkey. *Journal of Education and Practice, 6*(29), 131-141.

LaRoche, S., Joncas, M., & Foy, P. (2016). Sample design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 3.1-3.37). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-3.html

Lassen, S. R., Steele, M. M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to academic achievement in an urban middle school. *Psychology in the School, 43*, 701-711.

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.

Liaw, A. & Wiener, M. C. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18-22.

Liu, M., Hsieh, P., Cho, Y., & Schallert, D. (2006). Middle school students' self-efficacy, attitudes, and achievement in a computer-enhanced problem-based learning environment. *Journal of Interactive Learning Research, 17*(3), 225-242.

Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(1), 14-23.

Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, *82*(3), 329-348.

Mancini, L., & Sacco, C. (2020). A comparison of regression tree-based features selection methods for the prediction of academic performances. In P. Falzetti (Ed.), *INVALSI data: Methodologies and results III seminar " INVALSI data: A research tool"* (pp. 74-90). Milano: Franco Angeli.

Marchant, G. J. (2015). How plausible is using averaged NAEP values to examine student achievement?. *Comprehensive Psychology, 4*(1), 1-4.

Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Palazzo, L., & Foy, P. (2016). Creating and interpreting the TIMSS 2015 contex qoustionnaire scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 15.1-15.312). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-15.html

Milborrow, S. (2021). Plot 'rpart' models: An enhanced version of 'plot.rpart'. R package version 4.1-15. Retrieved from https://cran.r-project.org/web/packages/rpart.plot/rpart.plot.pdf

Ministry of National Education. (2016). *TIMSS 2015 national mathematics and science preliminary report.* Ankara: MEB. Retrieved from http://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_06/23161945_timss_2015_on_raporu.pdf

Mohtar, L. E., Halim, L., Samsudin, M. A., & Ismail, M. E. (2019). Non-cognitive factors influencing science achievement in Malaysia and Japan: An analysis of TIMSS 2015. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(4). doi:10.29333/ejmste/103567

Moisen, G. G. (2008). Classification and regression trees. In S. E. Jorgensen & B. D. Fath (Eds), *Encyclopedia of ecology* (pp. 582-588). Oxford, UK: Elsevier.

Motlagh, S. E., Amrai, K., Yazdani, M. J., Abderahim, H. A., & Sourie, H. (2011). The relationship between self-efficacy and academic achievement in high school students. *Procedia - Social and Behavioral Sciences, 15*, 765-768.

OECD. (2009). *PISA data analysis manuel: SPSS* (2nd ed.). Paris: OECD Publishing.

Osher, D., Bear, G. B., Sprague, J. R., & Doyle, W. (2010). How can we improve school discipline?. *Educational Researcher*, *39*(1), 48-58.

Ötken, Ş. (2019). Öğrencilerin fen başarılarını etkileyen değişkenlerin yapısal eşitlik modeli ile belirlenmesi: PISA 2015 Türkiye örneklemi. In H. Şahin (Ed.), *VI. Uluslararası Eğitim ve Sosyal Bilimlerde Akademik Çalışmalar Sempozyumu tam metin bildiri kitabı* (pp. 212-226). Elazığ: Asos Yayınevi.

Özer, A., Totan, T., & Atik, G. (2011). Individual correlates of bullying behaviour in Turkish middle schools. *Australian Journal of Guidance and Counselling, 21*(2), 186-202.

Pahmi, S., Saepudin, S., Maesarah, N., Solehudin, U. I., & Wulandari. (2018). Implementation of CART (classification and regression trees) algorithm for determining factors affecting employee performance. In *2018 International Conference on Computing, Engineering, and Design (ICCED) 6-8 September 2018* (pp. 57-62). Bangkok: Institute of Electrical and Electronics Engineers (IEEE).

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., EISPACK, Heisterkamp, S., … R-core. (2021). Linear and nonlinear mixed effects models. R package version 3.1-153. Retrieved from https://CRAN.R-project.org/package=nlme

Rutkowski, L., Gonzalez, E., Joncas, M., & Von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcers, 39*(2), 142-151.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods on Medical Resarch, 8*(1), 3-15.

Sela, R. J., & Simonoff, J. S. (2012). RE-EM trees: A data mining approach for longitudinal and clustered data. *Machine Learning, 86*(2), 169-207.

Sela, R. J., Simonoff, J., & Jing, W. (2021). Regression trees with random effects for longitudinal (panel) data. R package version 0.90.4. Retrieved from https://rdrr.io/cran/REEMtree/

Singh, K., Graville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interests and academic engagement. *Journal of Educational Research*, *95*(6), 323-332.

Skiba, R., & Peterson, R. (2003). Teaching the social curriculum: School discipline as instruction. *Preventing School Failure: Alternative Education for Children and Youth*, *47*(2), 66-73.

Speybroeck, N. (2012). Classification and regression trees. *International Journal of Public Health*, *57*(1), 243-246.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). New Jersey: Lawrance Erlbaum Association.

Tanaka, A., & Yamauchi, H. (2000). Causal models of achievement motive, goal orientation, intrinsic interest and academic achievement in classroom. *The Japanese Journal of Psychology*, *71*(4), 317-324.

Tezbaşaran, A. A. (2008). *Likert tipi ölçek hazırlama kılavuzu*. Retrieved from https://www.academia.edu/1288035/Likert_Tipi_%C3%96l%C3%A7ek_Haz%C4%B1rlama_K%C4%B1lavuzu

Therneau, T., Atkinson, B., & Ripley B. (2019). Recursive partitioning and regression trees. R package version 4.1-15. Retrieved from https://cran.r-project.org/web/packages/rpart/index.html

TIMSS & PIRLS. (2019). TIMSS 2015 international database. Retrieved from https://timssandpirls.bc.edu/timss2015/international-database/

Timofeev, R. (2004). *Classification and regression trees (CART) theory and applications* (Master's thesis). Humboldt University, Berlin.

Topçu, M. S., Erbilgin, E., & Arıkan, S. (2016). Factors predicting Turkish and Korean students' science and mathematics achievement in TIMSS 2011. *Eurasia Journal of Mathematics, Science & Technology Education*, *12*(7), 1711-1737.

Totura, C. M. W., Green, A. E., Karver, M. S., & Gesten, E. L. (2009). Multiple informants in the assessment of psychological, behavioral and academic correlates of bullying and victimization in middle school. *Journel of Adolescence, 32*(2), 193-211.

Trendowicz, A., & Jeffery, R. (2014). *Software project effort estimation: Foundations and best practice guidelines for success.* Switzerland: Springer International Publishing.

Uzun, N. B., Gelbal, S., & Öğretmen, T. (2010). TIMSS-R fen başarısı ve duyuşsal özellikler arasındaki ilişkinin modellenmesi ve modelin cinsiyetler bakımından karşılaştırılması. *Kastamonu Eğitim Dergisi*, *18*(2), 531-544.

Üstün, U., Özdemir, E., Cansız, M., & Cansız, N. (2020). Türkiye'deki öğrencilerin fen okuryazarlığını etkileyen faktörler nelerdir? PISA 2015 verisine dayalı bir hiyerarşik doğrusal modelleme çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *35*(3), 720-732.

van der Werf, C. (2014). The effects of bullying on academic achievement. *Revista Desarrollo y Sociedad, 74*, 275-308.

Whisman, A., & Hammer, P. C. (2014). *The association between school discipline and mathematics performance: A case for positive discipline approaches.* Charleston, WV: West Virginia Department of Education, Division of Teaching and Learning, Office of Research. Retrieved from https://files.eric.ed.gov/fulltext/ED569903.pdf

Wiberg, M., & Rolfsman, E. (2019). The association between science achievement measures in schools and TIMSS science achievements in Sweden. *International Journal of Science Education, 41*(16), 2218-2232.

Yamauchi, K., Ono Y., Baba, K., & Ikegami, N. (2001). The actual process of rating the global assesment of functioning scale. *Comprehensive Psychiatry, 42*(5), 403-409.