Psychological Test and Assessment Modeling, Volume 61, 2019 (4), 453-470

Using tree-based regression to examine factors related to math ability among 15-year old students

Cody Ding[1,2] & Yuyang Zhao[2]

Abstract

Few studies have examined factors that can predict students' mathematic ability, particularly subgroups of students who share the common characteristics that are associated with different levels of math ability. Based on PISA 2012 data from the United States and China, we used regression tree analysis to select the most salient predictors of math ability and identify the subgroups of 15-year-old students who were likely to be proficient in math ability. Based on the results from regression tree analysis, it was found that students whose math self-efficacy score was over 3.33 and their perceived positive peer math norm score was below 3.33 on a rating scale of 1 to 4 were most likely to be associated with proficient math ability. In contrast, students whose math self-efficacy score was below 2.81 were most likely to be associated with low or below-average math ability. However, for students whose math self-efficacy score was between 2.81 and 3.33, their math ability level was likely to be associated with their perceived positive peer math norm. The significance of the study is that it uniquely identified distinct subgroups of students who were more likely to be associated with different levels of math ability. Methodologically, the study demonstrated the application of regression tree analysis in studying students' math ability.

Keywords: mathematic ability, regression tree, tress-based method, math self-efficacy, PISA 2012

[1] *Correspondence concerning this article should be addressed to:* Cody Ding, PhD, Education Sciences & Professional Programs, University of Missouri-St. Louis, St. Louis, USA: email: dingc@umsl.edu

[2] University of Missouri-St. Louis, USA

Students' math achievement has been an enduring research topic for many years. Researchers and practitioners have been interested in how variables of parents, schools, and students are associated with math achievement. For example, using regression analysis, Areepattamannil, Khine, Melkonian, Welch, Al Nuaimi, and Rashad (2015) found that students who believed their parents had a positive attitude toward mathematics performed significantly better on mathematic assessments compared to students whose parents did not focus on learning math. Similarly, students' attitude toward mathematics, such as math self-concept, math self-efficacy, and math interest were found to be positively related to math achievement (e.g., Arens et al., 2016; Catapano, 2014; Ma & Kishor, 1997; Marsh & Hau, 2004; Sharma, 2013; Yüksel & Geban, 2016). Among all the student-related variables, mathematics self-efficacy was found to be the most salient factor regarding mathematics achievement ($0.42 \leq r \leq 0.57$) (Bonne & Johnston, 2016; Kalaycioglu, 2015).

One noticeable gap in the math literature was the lack of discussion regarding the difference between math achievement and math ability or skills. Previous literature has mainly focused on mathematic achievement; that is, what has been taught and learned at school rather than the math ability that reflects and applies what has been taught (i.e., not curriculum-based). A significant difference between the two constructs (i.e. math achievement vs. math ability or skills) was articulated by the Organization for Economic Cooperation and Development (OECD) (2013): "An individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognize the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens" (p. 26). Accordingly, the Programme of International Student Assessment (PISA) developed ingenious ways to assess math ability. Specifically, math ability is conceptualized as the ability to reflect on knowledge and experience, and to apply that knowledge and experience to real world issues (p. 14) (Program for International Student Assessment, 2001). Thus, math ability is not to assess what skills schools require students to learn but what skills society requires of young people in the workforce. In a sense, math ability is a forward-looking construct that emphasizes students' ability to use knowledge and skills learned in school to meet real-life challenges, rather than the extent to which they have mastered school math curriculum (Program for International Student Assessment, 2001).

This difference between math achievement and math ability may be supported by a study conducted in Germany, which found that the state-mandated graduation exams for German secondary schools improved their students' curriculum-based knowledge, but not their mathematical ability or skills (Jürges, Schneider, Senkbeil, & Carstensen, 2012). The finding seemed to suggest the key difference between math achievement (i.e., curriculum-based knowledge as measured by achievement tests) and math ability (i.e., application of what is learned in school). Thus, students might be able to obtain a relatively high test-score or learn a specific curriculum-based knowledge concept very well, but they might not be able to apply or generalize that concept to real-world problems. Therefore, the question is: what factors predict students' math ability?

Reviewing literature on student mathematical learning did not yield many articles on the topic of mathematical ability or skills as defined in PISA. Given what we have already learned about the factors related to math achievement (e.g., math self-concept, math anxiety, peer norm, etc.), we examined in the current study whether the student-related factors were also relevant to student mathematics ability or skills. Culling from the literature regarding student math achievement, we hypothesized that the factors (e.g., math self-efficacy) related to the students' math achievement would have similar effects on the students' math ability. Moreover, we wanted to further identify distinct subgroups of students who shared the common characteristics that were likely to be related to a particular level of math ability.

Method

Data source

The data used in the current study were from the Programme for International Student Assessment (PISA) 2012, in which 34 OECD member countries as well as 31 non-member partner countries and economies participated. In the present study, we only used the data from the US (n = 5,140) and China (n = 5.331). The participants' ages ranged from 16 years to 17 years. The sample included 5,234 girls (50%), and ethnicity information was not provided.

Measures

All measures used were based on items from PISA 2012. In the following, we discussed each measure in detail.

Math ability. Based on the information provided by PISA 2012 (OECD, 2014), math ability was assessed using two formats: paper-based instruments and optional computer-based components (CBAM). The math assessment contained a total of 108 items with 72 new items and 36 old items from earlier surveys. Students were given a total of two hours to complete the paper-based assessment (an additional 40 minutes were devoted to CBAM). The paper-based instruments contained a total of 270 minutes of mathematics material, which was arranged in nine clusters of items. Each participating country used seven clusters: three link materials, two new 'standard' clusters, and either the two 'standard' or two 'easy' clusters.

Three types of question formats were included: open constructed-response, closed constructed-response, and selected-response (multiple-choice). Student responses were either manually coded by trained experts or automatically coded by data capture software. The CBAM contained a total of 80 minutes of mathematics material, which was arranged in four clusters of items. Additional format types were possible in CBAM because of the computer-based environment, such as the manipulation and rotation of representations of three-dimensional shapes. Calculators or other equivalent functionality tools were permitted for use in both paper-based instruments and CBAM. More detailed information can be found in the PISA 2012 framework (OECD, 2014).

The items of the PISA 2012 mathematics assessment focused on four content categories – quantity, uncertainty and data, change and relationships, and space and shape. We used the average score of these four areas as our measure of mathematical ability. The mean score of each content area was provided in the data. The mean scores ranged from 0.48 to 0.86, with a higher mean score indicating a higher level of math ability. For the purpose of tree-based regression analysis (discussed below), we categorized math ability scores into four levels: low, below average, above average, and proficient.

Math self-efficacy. Eight items were used to measure students' math self-efficacy (e.g., calculating TV discount, understanding graphs in newspapers, and distance to scale) on a 4-point rating scale, with *1* indicating "Very confident" and *4* indicating "Not at all confident." These eight items formed one factor based on our results of exploratory factor analysis (EFA), with factor loading ranging from 0.69 to 0.82 and a median of 0.74. The Cronbach's Alpha was 0.89. The scale of each item was recoded and a mean score was created, with a higher mean score indicating a higher level of math self-efficacy.

Student perceived peer math norms. Perceived peer math norm was defined in this study as how students perceived peers' math performance. There were three items used to evaluate students' perception about math norms (e.g., most of my friends do well in mathematics, most of my friends work hard at mathematics, and my friends enjoy taking mathematics tests) on a 4-point rating scale, with *1* indicating "Strongly agree" and *4* indicating "Strongly disagree." The results from our EFA on these three items showed factor loading ranging from 0.73 to 0.81, with a median of 0.80. The Cronbach's Alpha was 0.85. The scale of each item was recoded and a mean score was created, with a higher mean score indicating a higher level of perceived positive peer math norms.

Math anxiety. Five items were used to measure students' math anxiety (e.g., worry that it will be difficult, get very tense, and feel helpless) on a 4-point rating scale, with *1* indicating "Strongly agree" and *4* indicating "Strongly disagree." Five items formed one factor based on the results of EFA, with factor loading ranging from 0.74 to 0.85 and a median of 0.82. The Cronbach's Alpha was 0.82. The scale of each item was recoded and a mean score was created, with a higher mean score indicating a higher level of math anxiety.

Math self-concept. There were five items used to assess students' math self-concept (e.g., not good at math, get good grades, and understand difficult mathematic work) on a 4-point rating scale, with *1* indicating "Strongly agree" and *4* indicating "Strongly disagree." We removed one negatively worded item because research indicated that negatively worded items have different wording effects on factor structure of items compared to positively worded items (e.g., Gu, Wen, & Fan, 2015). The remaining four items formed one factor based on the results of EFA, with factor loading ranging from 0.79 to 0.86 and a median of 0.84. The Cronbach's Alpha was 0.86. The scale of each item was recoded and a mean score was created, with a higher mean score indicating a higher level of math self-efficacy.

Math work ethic. There were 9 items used to evaluate students' math work ethic (e.g., homework completed in time, prepared for exams, pay attention in class, and listen in class) on a 4-point rating scale, with *1* indicating "Strongly agree" and *4* indicating "Strongly disagree." The results from EFA on these 9 items showed two factors – students' after-class math work ethic (5 items), with factor loading ranging from 0.57 to 0.88 and a median of 0.77, and students' in-class math work ethic (4 items), with factor loading ranging from 0.41 to 0.98 and a median of 0.80. The Cronbach's Alpha was 0.77 for after-class math work ethic and 0.84 for in-class math work ethic, respectively. The scale of each item was recoded and a mean score was created, with a higher mean score indicating a higher level of after-class or in-class math work ethic.

Data analysis

For the purpose of identifying variables that are related to math ability, the most common method is regression analysis. Due to advances in computer technologies, there are more types of regression analytical methods available now than ever that can be routinely used, including logistic regression, robust regression, quantile regression, or tree-based regression, to name a few. In the current study, we could use ordinal logistic regression analysis since our response variable of math ability was ordered categorical variable. One of the assumptions of ordinal logistic regression is proportional odds assumption, which indicates that the slope estimate for each pair of groups should be the same across all pair groups. In this case, the χ^2 test for proportional odds assumption (i.e., slope coefficients are the same across response categories) was 20.88, $p = .007$, indicating that the cumulative logit model might not adequately fit the data. In addition, we wanted to identify subgroups in the sample that shared the common characteristics that is most likely to be related to different types of math ability.

Tree-based regression. An alternative method can be tree-based regression analysis. Regression tree can fit many kinds of statistical model, including least-square, quantile, logistic, and Poisson (Loh, 2014). Since it is a tree-based method, regression tree is non-parametric and robust to violation of assumptions related to regular regression analysis. It can also incorporate missing data in the analysis without need of imputation. In addition, it provides a visual output that has a multilevel structure, resembling an inverse tree structure. This tree structure provides best classification of subgroups based on optimal cutoff score when the predictors are continuous variables.

Given the greater availability of software (e.g., SAS, R, or SPSS), tree-based regression analysis can easily be accessible, gaining acceptance and popularity in social science research. In the following sections, we described tree-based regression in a bit more detail, given its unfamiliarity to some researchers. There are many books, articles, and computer program documents on this topic. Materials presented here are mainly based on those from SAS documents for tree-based analysis due to its readability (SAS, 2013).

Regression tree may be viewed as a variant of decision tree methodology that was first developed by Breiman and colleagues (Breiman, Friedman, Olshen, & Stone, 1984), which is designed to efficiently estimate meaningful subgroups. That is, it allows us to identify subgroups of individuals who are most likely to be associated with a particular behavioral outcome. The subgroups identified are mutually exclusive and exhaustive, and they share similar characteristics with the response variable of interest. Usually, regression trees start with a response variable or a node that contains all cases, called parent node. Then, all predictors are examined in order to select the variables that results in binary groups that are most predictive with respect to the response variable. Subsequently, the predictor selected is split into two child nodes. Each response variable or the node only divides into two child nodes. Within each of these two child nodes, the tree-growing method continues to further split or subdivide each child node into more binary groups by assessing the importance of the remaining predictors. Thus, each child node becomes a parent node with respect to subsequent predictors.

Typically, regression trees are produced by algorithms that split all possible combinations of values of predictors into binary groups or regions, also known as leaves. Although there are several variations of tree-splitting algorithms, the basic one is automatic interaction detection (AID) (Morgan & Sonquist, 1963). As Loh (2014) indicated, AID splits the values in the parent node into two children by recursive partitioning, starting with all values at the top of the tree (i.e., parent node). The algorithm splits this parent node into two or more child nodes in such a way that the responses within each child region are as similar as possible. At each step, the split is determined by finding a best predictor and a best cutoff point that assigns the cases in the parent node to the child nodes. The splitting process is then repeated for each of the child nodes, and the recursive partitioning process continues until a stopping criterion is satisfied and the tree is fully built. The nodes without any child nodes are terminal nodes.

There are two types of criteria for splitting a parent node: criteria that minimize node impurity and criteria that is defined by a statistical test. Minimizing node impurity is defined by

$$\phi_{(t)} = i_{(t)} - \sum_{c=1}^{C} p\left(t_c|t\right) i_{t_c} \tag{1}$$

where $\phi_{(t)}$ is the impurity of node t and t_c is the c^{th} child node, $p\left(t_c|t\right)$ is the proportion of cases in node t that are assigned to t_c, and C is the number of branches after splitting parent node t. The criteria based on impurity reduction criteria employ different impurity functions $i_{(t)}$, which include the entropy, Gini index (Hastie, Tibshirani, & Friedman, 2009), and residual sum of square (RSS).

Criteria based on statistical test include the χ^2, F test, CHAID (chi-square AID), or FastCHAID, all of which compute the worth of a split by testing significant differences in the response variable across the branches defined by a split. The worth of a split is defined by -log(p), where p is the p-value of the test. Typically, an F test is used with a continuous response variable.

When building a tree, it is computationally intensive to calculate the node purity or the node worth for all possible cutoff points of every predictor. Thus, there are different splitting strategies to generate possible splits. These strategies include the exhaustive method, which computes the split criterion for all the levels of a predictor, the greedy method, which finds the split by recursively halving the data based on CHAID algorithm, and the fast-sort method, which prioritizes the split according to the cases of levels of a categorical response or the average of continuous response.

However, the splitting strategies described above can result in potentially overfitting the training data and not generalizing to new data. Accordingly, it is necessary to prune the tree to a smaller subtree in order to lower error rate on validation data. This is done by focusing on a sequence of nested trees obtained by successively pruning leaves from the tree. There are several pruning methods, including cost-complexity pruning (Breiman et al., 1984), C4.5 pruning (Quinlan, 1993) and reduced-error pruning (Quinlan, 1987), which is based on the average of square error (ASE).

After tree building, we can assess model fit to the data. Various measures of model fit have been proposed in the datamining literature. These model fit indices include entropy, Gini index, misclassification rate, RSS, ASE, sensitivity or area under the curve (AUC) (for binary response variable), and confusion matrix. In addition, the importance of each predictor can also be assessed; tree-based regression typically selects the most useful predictors based on an indication of which predictors are most useful in predicting the response variable. The indication of variable importance includes, (1) count, which simply counts the number of times in the tree that a particular predictor is used in a split, (2) surrogate count, which tallies the number of times that a predictor is used in a surrogate splitting rule, (3) RSS, which is based on change of RSS when a split is found at a node, and (4) relative importance, which is a number between 0 and 1, with 1 indicating the most important predictor.

In the current study, the regression tree was built with math ability as a response variable. The predictors included math self-efficacy, math work ethic, math anxiety, student perceived math-norm, and math self-concept. It should be noted that the correlation between math self-efficacy and math self-concept was .38 based on the sample used, although conceptually these two variables should be highly correlated. One reason was that math self-efficacy was operationalized as being able to perform a math calculation in PISA 2012, whereas math self-concept was operationalized as self-perceived competence in learning (e.g., learn quickly or get good at). Thus, it was possible that the students had a high self-perception but was not able to perform or vice versa, resulting in a low correlation between the two variables.

As mentioned previously, math ability was treated as an ordered categorical response variable, and the predictors were all treated as continuous variables. In building the regression tree, we used entropy as the splitting criteria and cost-complexity as the pruning method. The selection of the pruning parameter was based on cross validation, in which random selection of the cases was used to obtain our validation sample. In the analysis, missing values were handled by listwise deletion because 66% of data on the response variable of math ability was missing.

K-fold cross validation is used as a model evaluate method. Specifically, we used 10-folds cross validation in which data was divided into 10 subsets and the simple holdout method (i.e., one subset was used as validation sample (10%) and remaining k – 1 subsets were used as training sample) was repeated 10 times. The average error or measure of accuracy across 10 folds was computed. In k-fold cross validation, every data point got to be in a validation sample exactly once and in a training data k – 1 times. The advantage of this method is that it did not matter much how the data were divided. In contrast, simple holdout method may lead model evaluation to be significantly different depending on how the data split was made in terms of which data points ended up in the training set and which ended up in the validation set.

Model fit was assessed by confusion matrix, ASE, misclassification, Gini, and RSS based on 10-fold cross validation method. The importance of predictors was assessed by relative importance and count. The analysis was performed using SAS 9.4 (SAS, 2013).

Results

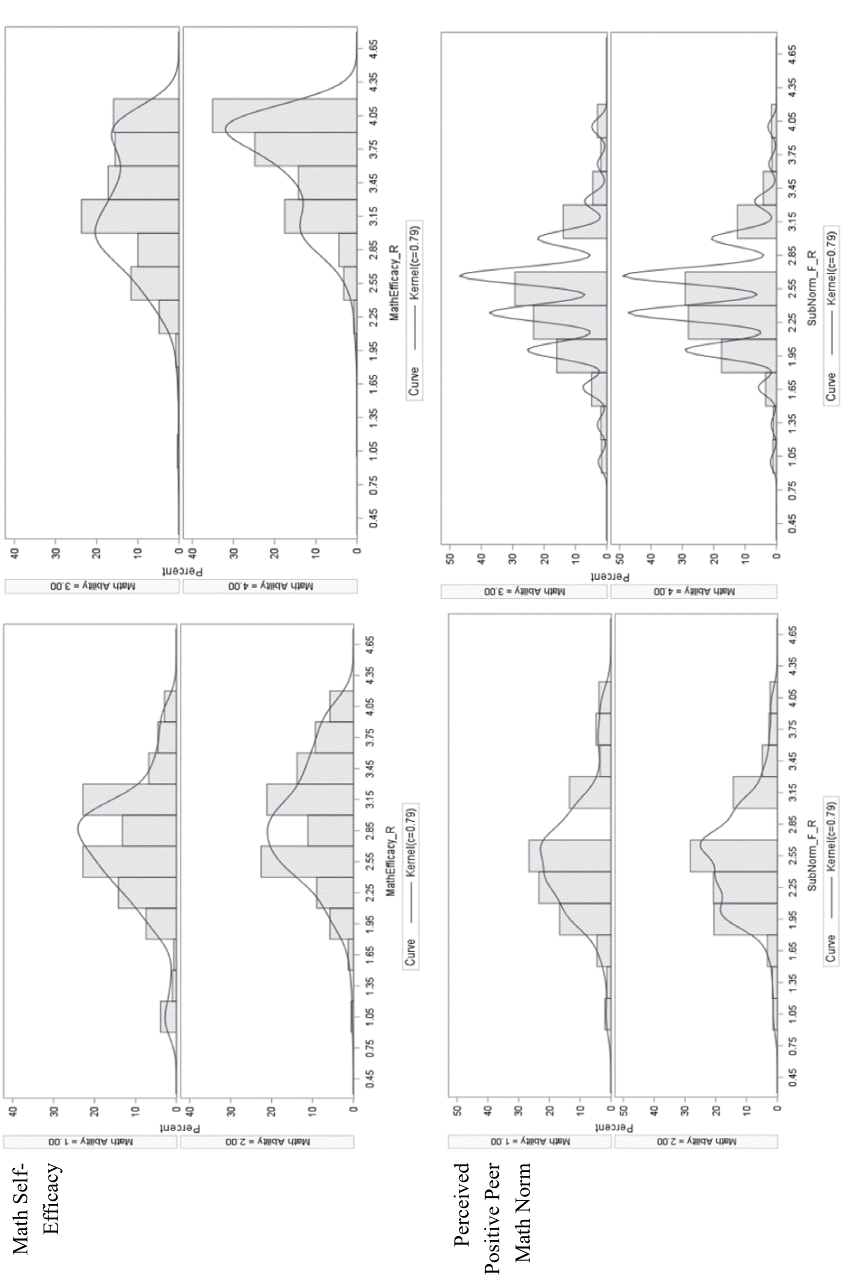
Figure 1 shows the kernel distribution of predictor variables by the four levels of math ability. Overall, the distribution of predictor variables was not markedly deviating from the normality at each level of math ability. Skewness across four levels of math ability ranged from -0.82 to 0.27, with the median being 0.75, and kurtosis ranged from -0.53 to 1.29, with the median being -0.67.

Figure 2 displays estimates of the validation misclassification rate for a series of progressively smaller subtrees of the large tree that are indexed by a cost-complexity parameter. It is also known as a pruning or tuning parameter. The figure provided a visual tool for selecting the parameter that yielded the smallest misclassification rate, which was indicated by the vertical dotted reference line. The subtree size (i.e., number of leaves) that corresponded to each parameter was indicated on the upper horizontal axis. The parameter value 0 corresponded to the fully-grown tree, which had 44 leaves. However, the misclassification rate for two smaller subtrees, one with nine leaves and one with twelve leaves, was indistinguishable from the minimum misclassification rate. This was evident from a comparison of the standard errors for the misclassification rate, as indicated by the error bar.

Table 1 shows the 10-fold cross-validation assessment of the model. As can be seen in the table, the subtree with eight leaves had the smallest ASE and an average misclassification rate. The results from validation assessment of the model seemed to be very similar to those from the training data. Thus, we selected a smaller tree with eight leaves or nodes to avoid overfitting. It should be noted that the estimated ASEs and their standard errors depend on the random allocation of the cases to the 10 folds of validation samples. Subsequently, we may have obtained different estimates if we used a different random seed value. Similarly, the estimates may have differed if we had used a different number of folds.

Table 2 shows the accuracy and the model fit of the selected tree model with eight nodes. Confusion matrix indicated that error rate was high for classification of categories Low and Above Average, suggesting that the model might not be accurate in predicting cases for these two categories. This result was based on both training data and validation samples. The model fit indices indicated over 50% misclassification rate, suggesting a bit more uncertainty in prediction using the selected regression tree model, as indicated by entropy value.

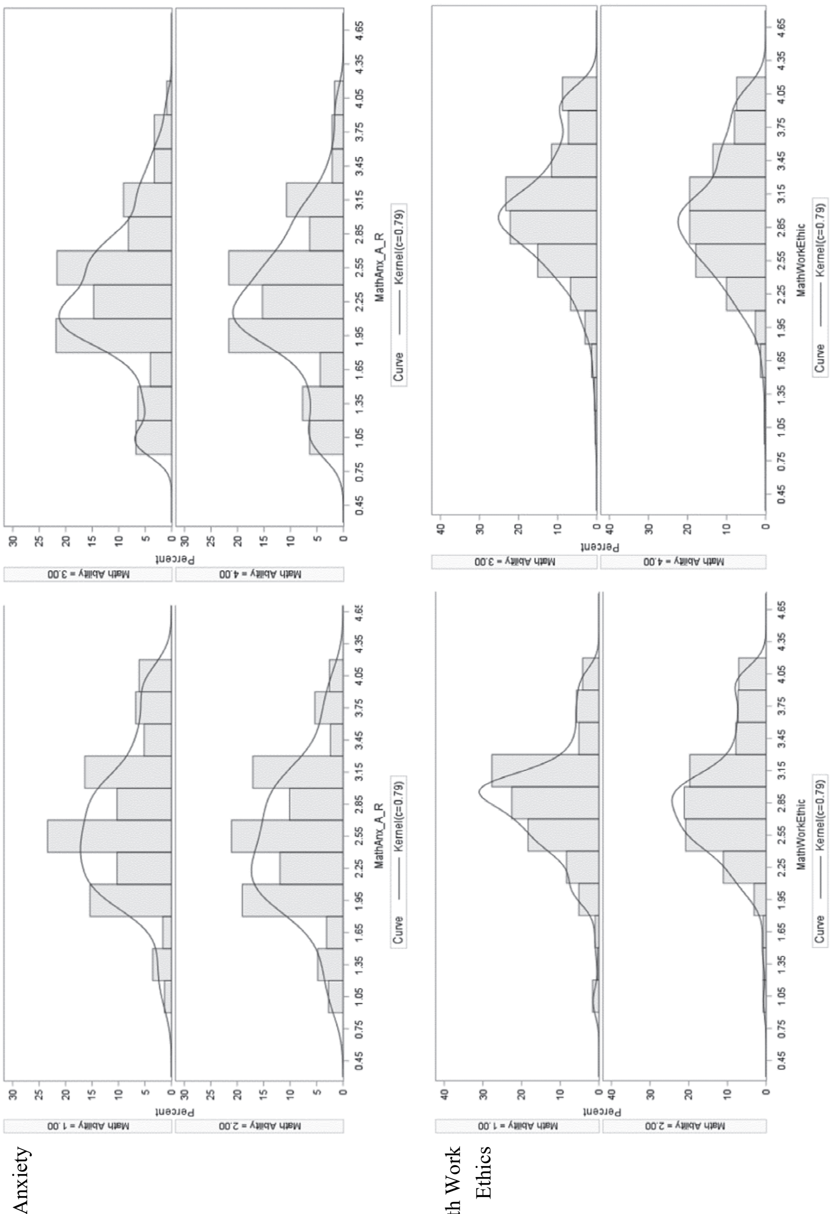
Figure 1:
Kernel distribution of predictor variables by level of math ability. Math ability 1 = low, 2 = below-average, 3 = average above, 4 = proficient



Continuation of Figure 1

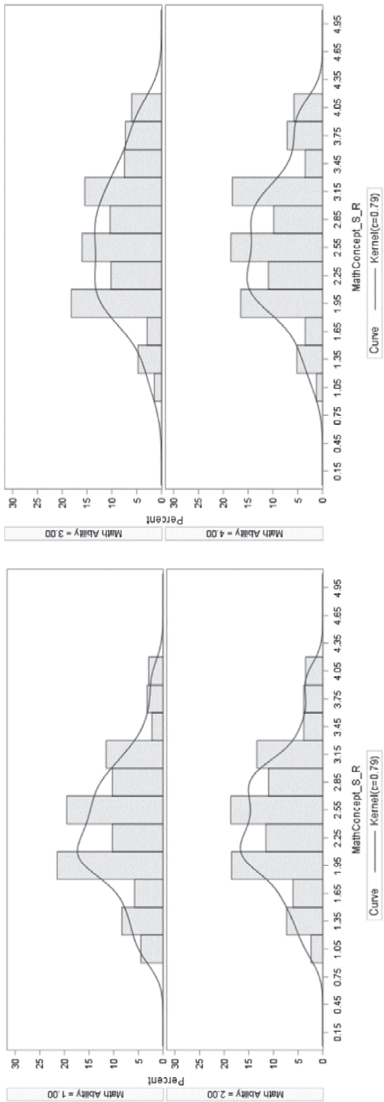
Math Anxiety

Math Work Ethics



Continuation of Figure 1

Math
Self Concept



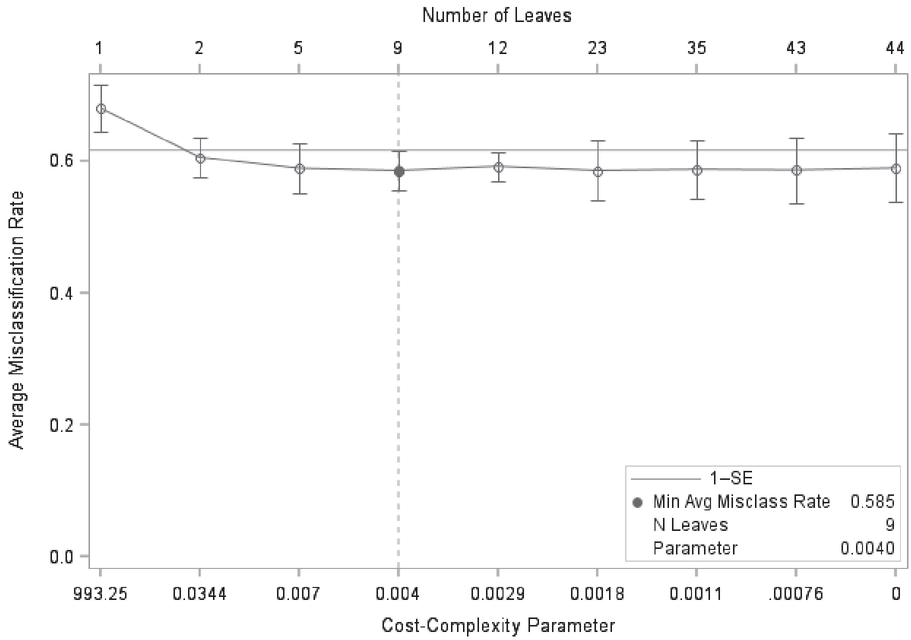


Figure 2
Validation misclassification as a function of cost-complexity parameter

Table 1:
Cross Validation Assessment of the Model

N Leaves	Average Square Error				Number of Leaves			Misclassification Rate			
	Min	Avg	Standard Error	Max	Min	Median	Max	Min	Avg	Standard Error	Max
8	0.163	0.172	0.006	0.181	6	8.0	15	0.522	0.589	0.055	0.687

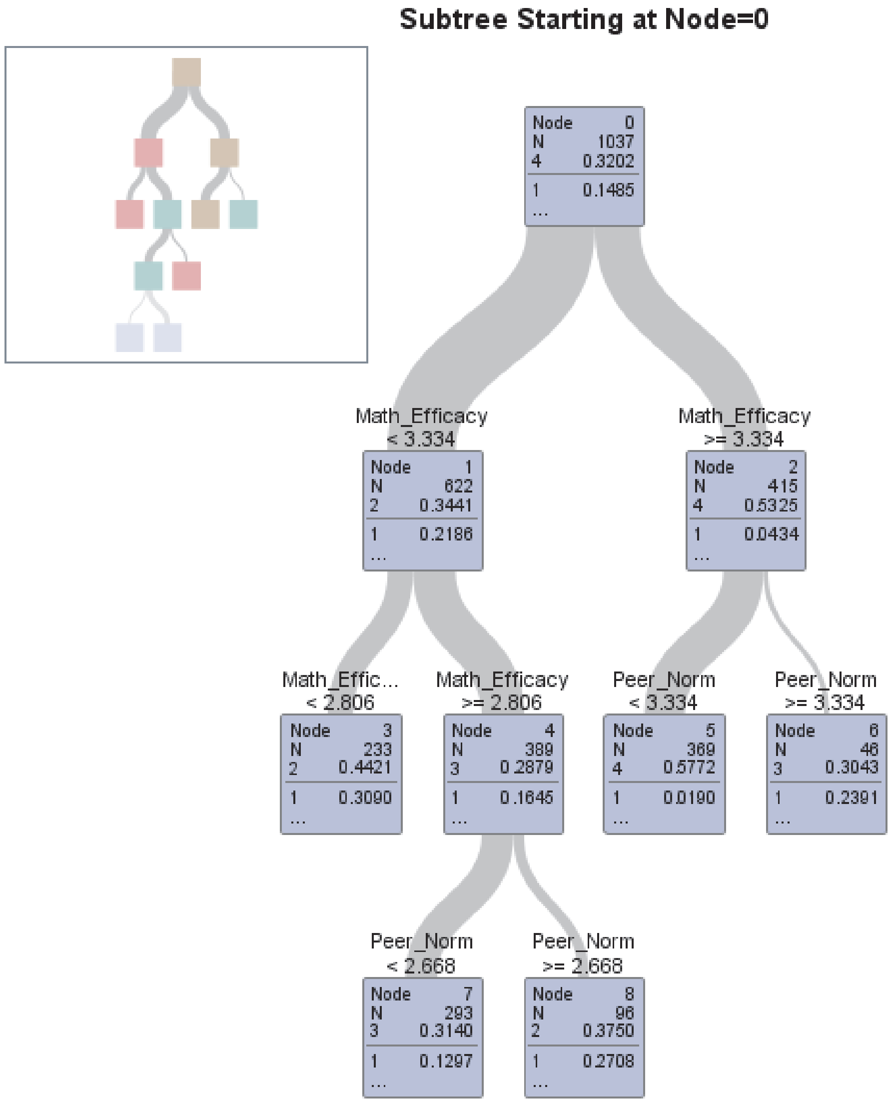
Table 2:
Confusion Matrix for Classification and Model Fit for Selected Regression Tree

Confusion Matrices						
	Actual	Predicted				Error Rate
		Low	Below Average	Above Average	Proficient	
Model Based	Low	0	121	15	18	1.000
	Below Average	0	188	28	64	0.328
	Above Average	0	111	47	113	0.826
	Proficient	0	57	25	250	0.247
Cross Validation	Low	9	88	31	26	0.941
	Below Average	14	128	59	79	0.542
	Above Average	3	81	65	122	0.760
	Proficient	3	43	63	223	0.328

Model Fit for Selected Tree						
	N Leaves	ASE	Misclassification Rate	Entropy	Gini	RSS
Model Based	8	0.163	0.532	1.694	0.652	675.8
Cross Validation	8	0.172	0.589			

Figure 3 shows the details about the final tree, including splitting variables and values. At the top of the regression tree, Node 0 indicated the response variable of math ability, which had four levels (i.e., low, below average, above average, and proficient), with category 4 (proficient) having the most cases and category 1 (low) having the least cases. The first split was based on the math self-efficacy variable. There were 622 cases with math self-efficacy scores less than 3.33 (Node 1) and 415 cases with math self-efficacy scores over 3.33 (Node 2). Node 1 was further split into two groups with math self-efficacy scores that were either less than 2.806 (Node 3) or over 2.806 (Node 4). No further split was noted for cases in Node 3, which was likely to be associated with below-average (44.2%) or low math ability (30.9%). In contrast, for cases in Node 4, perceived positive peer math norm provided the most significant split (Nodes 7 and 8). Cases in Node 7 were likely to be associated with above-average math ability (31.4%), while cases in Node 8 were likely to be associated with below-average math ability (37.5%) or low math ability (23.9%).

Among those whose math self-efficacy score was over 3.33 (Node 2), perceived positive peer math norm provided the most significant split at a score cutoff point of 3.33 (Nodes 5 and 6). No further splits were observed for cases in Nodes 5 and 6. Cases in Node 5 were likely to be associated with proficient math ability (58%), and cases in Node 6 were likely to be associated with low math ability (23.9%) or above-average math ability (30.4%). It should be noted that Nodes 3, 5, 6, 7, and 8 were terminal nodes or leaves since they had no child nodes.



1 = Low 2 = Below Average 3 = Above Average 4 = Proficient

Figure 3:
Final regression tree model

Based on these results, it seemed that proficient math ability was most likely to be associated with those whose math self-efficacy score was over 3.33 and perceived positive peer math norm was below 3.33. However, individuals with low math self-efficacy scores (< 2.806) were likely to have low or below-average math ability. This was true for those whose perceived positive peer math norm score was over 2.806.

Table 3 shows the importance of the variable in predicting math ability. The result indicated that math self-efficacy seemed to be the most important predictor, followed by perceived peer math norm. Math anxiety, math self-concept, and math work ethics were not included in the model as the best predictors.

Table 3:
Variable Importance

	Relative Importance		Count
Math Efficacy	1.00	7.88	2
Peer Math Norm	0.48	3.80	2
Math Work Ethic	0.30	2.38	1

Discussion

The purpose of the current study was to examine the factors related to math ability rather than math achievement, using tree-based regression based on data from PISA 2012. Math ability was defined as the ability of 15-year-old students to apply math knowledge they learned from school to real-world math problems with respect to quantity, uncertainty and data, change and relationships, and space and shape. The key finding was that math self-efficacy was the most salient predictor of math ability. Perceived positive peer math norm (i.e., if their peers were doing well in math or not) was also associated with math ability. These findings were consistent with previous studies that showed math self-efficacy and classmate characteristics were associated with students' learning outcomes (e.g., Chiu & Chow, 2015).

What was unique about the findings from the regression tree was that it found the best score cutoff point in predictors that may be most likely to be associated with a particular category of the outcome. That is, it has an ability to easily segment cases into distinct subgroups whose members share similar characteristics related to a particular behavior. In the current case, the score cutoff point of 3.33 for both math self-efficacy and perceived positive peer math norm was noticeably associated with proficient math ability. Since these two predictors were measured on a scale of 1 to 4, this cutoff point of 3.33 indicated that students' math self-efficacy scores needed to be above the 83rd percentile and their perceived peer math norm score needed to be below the 83rd percentile for them to be math skill proficient. In contrast, regular regression models are designed to determine the average effect of an independent variable on a response variable and consider the average member of the population rather than considering special needs or characteristics of subgroups (Forthofer & Bryant, 2000).

It is worth mentioning the relationships between perceived positive peer math norms and proficient math ability. Students who had a lower level of perceived positive peer math norm tended to have a higher level of math ability. One possible reason may be that students who had proficient math ability may feel that their peers may not do as well in math as themselves. For those students who had a lower level of math ability, they may feel that their peers do better in math than themselves. It may be through this sign of comparison or reference point that perceived positive peer math norm played a role in predicting the level of students' math ability.

One implication of knowing the score cutoff point that is likely to be associated with certain outcomes is that it may help us to better identify the target level of intervention efforts. For example, in order for students to be math skill proficient, we need to improve their math self-efficacy score to be at the top 17th percentile. This explicit target may help teachers or parents set a clear goal and action plan. Similarly, in psychological assessment, identifying a score cutoff point is necessary for clinical diagnosis, and tree-based regression may be a better way to validate the cutoff point for the purpose of classification, particularly when there is a need for identifying at-risk populations.

There are limitations related to the current study and the regression tree. First, the sample may not be representative of the population of 15-year-old students from both US and China because the participants of the PISA study were from selected regions of the country. In the sample from China, for example, these participants were mainly from Shanghai, which tends to show a higher level of education quality as well as economic conditions. Second, there are various tree-growing methodologies, such as CHAID, CART, or Gini. The different splitting criteria or stopping rules may well result in different conclusions, which makes it harder to decide which one can better approximate the reality. That is, sometimes regression trees can become complex, growing into multiple levels and splits that have no practical significance and are difficult to interpret. Third, tree-based methods are often used in data mining, in which data snooping happens when all possible predictors are entered to see what results will occur. This may be particularly true in today's technology climate when big data analysis or artificial intelligence are made possible without a priori consideration of theory or expectations. Therefore, it is important more than ever to develop specific hypotheses based on theories or previous findings.

Despite these limitations, tree-based regression still provides a viable methodology through which we can identify distinct subgroups that share the common characteristics related to a particular outcome, as we did in this study.

References

Areepattamannil, S., Khine, M. S., Melkonian, M., Welch, A. G., Al Nuaimi, S. A., & Rashad, F. F. (2015). International note: Are Emirati parents' attitudes toward mathematics linked to their adolescent children's attitudes toward mathematics and mathematics achievement? *Journal of Adolescence, 44*(1), 17-20.

Arens, A. K., Marsh, H. W., Craven, R. G., Yeung, A. S., Randhawa, E., & Hasselhorn, M. (2016). Math self-concept in preschool children: Structure, achievement relations, and generalizability across gender. *Early Childhood Research Quarterly, 36*, 391-403. doi: 10.1016/j.ecresq.2015.12.024

Bonne, L., & Johnston, M. (2016). Students' beliefs about themselves as mathematics learners. *Thinking Skills and Creativity, 20*, 17-28. doi: 10.1016/j.tsc.2016.02.001

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Catapano, M. (2014). *Tenth-grade high school students' mathematical self-efficacy, mathematics anxiety, attitudes toward mathematics, and performance on the New York state integrated algebra regents examination.* (74), ProQuest Information & Learning. Retrieved from https://ezproxy.umsl.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-99111-230&site=ehost-live&scope=site Available from EBSCOhost psyh database.

Chiu, M. M., & Chow, B. W. Y. (2015). Classmate characteristics and student achievement in 33 countries: Classmates' past achievement, family socioeconomic status, educational resources, and attitudes toward reading. *Journal of Educational Psychology, 107*(1), 152-169.

Forthofer, M. S., & Bryant, C. A. (2000). Using audience-segmentation techniques to tailor health behavior change strategies. *American Journal of Health Behavior, 24*, 36-43.

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2 ed.). New York: Springer-Verlag.

Jürges, H., Schneider, K., Senkbeil, M., & Carstensen, C. H. (2012). Assessment drives learning: The effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review, 31*(1), 56-65. doi: 10.1016/j.econedurev.2011.08.007

Kalaycioglu, D. B. (2015). The Influence of Socioeconomic Status, Self-Efficacy, and Anxiety on Mathematics Achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and the USA. *Educational Sciences: Theory and Practice, 15*(5), 1391-1401.

Loh, W.-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review, 82*(3), 329-348. doi: 10.1111/insr.12016

Ma, X., & Kishor, N. (1997). Assessing the Relationship between Attitude toward Mathematics and Achievement in Mathematics: A Meta-Analysis (Vol. 28, pp. 26-47): Journal for Research in Mathematics Education.

Marsh, H. W., & Hau, K.-T. (2004). Explaining Paradoxical Relations Between Academic Self-Concepts and Achievements: Cross-Cultural Generalizability of the Internal/External

Frame of Reference Predictions Across 26 Countries. *Journal of Educational Psychology, 96*(1), 56-67. doi: 10.1037/0022-0663.96.1.56

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of American Statistical Association,, 58*, 415-434.

OECD. (2014). PISA 2012 Technical report. Paris: OECD Publishing.

Organization for Economic Co-operation, & Development. (2013). PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy. from OECD Publishing http://dx.doi.org/10.1787/9789264190511-en

Program for International Student Assessment. (2001). Knowledge and skills for life: First results from PISA 2000 (pp. 14). Paris: OECD Publishing.

Quinlan, J. R. (1987). "Simplifying Decision Trees." *International Journal of Man-Machine Studies, 27*, 221-234.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann.

SAS. (2013). *Base SAS® 9.4 procedures guide: Statistical procedures* (2 ed.). Cary, NC: SAS Institute, Inc.

Sharma, A. (2013). *Associations between Self-Efficacy Beliefs, Self-Regulated Learning Strategies, and Students' Performance on Model-Eliciting Tasks: An Examination of Direct and Indirect Effects.* ProQuest LLC. Retrieved from https://ezproxy.umsl.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED558272&site=ehost-live&scope=site

http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqm&rft_dat=xri:pqdiss:3586451 Available from EBSCOhost eric database.

Yüksel, M., & Geban, Ö. (2016). Examination of Science and Math Course Achievements of Vocational High School Students in the Scope of Self-Efficacy and Anxiety. *Journal of Education and Training Studies, 4*(1), 88-100.