

RESEARCH

Open Access



# Machine learning procedures for predictor variable selection for schoolwork-related anxiety: evidence from PISA 2015 mathematics, reading, and science assessments

Jason C. Immekus<sup>1</sup>, Tai-sun Jeong<sup>2</sup> and Jin Eun Yoo<sup>3\*</sup> 

\*Correspondence:  
jineun.yoo@gmail.com

<sup>1</sup> Department of Educational Leadership, Evaluation, and Organizational Development, University of Louisville, 1905 S. 1st St., Louisville, KY 40292, USA

<sup>2</sup> Korea Institute of Curriculum and Evaluation, 8, Gyohak-ro, Deoksan-eup, Jincheon-gun, Chungcheongbuk-do 27873, Republic of Korea

<sup>3</sup> Department of Education, Korea National University of Education, 250 Taesungbyeon-ro, Ganenae-Myeon, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28713, Republic of Korea

## Abstract

**Background:** Large-scale international studies offer researchers a rich source of data to examine the relationship among variables. Machine learning embodies a range of flexible statistical procedures to identify key indicators of a response variable among a collection of hundreds or even thousands of potential predictor variables. Among these, penalized regression approaches, including least absolute selection and shrinkage operator (LASSO) and elastic net (Enet), have been advanced as useful tools capable of handling large number of predictors for variable selection for model generation. While the utility of penalized regression within educational research is emerging, less application of these machine learning methods, including random forest, to predictor variable selection in large-scale international data appears in the literature. In response, this study compared LASSO, Enet, and random forest for predictor variable selection, including the traditional forward stepwise (FS) regression approach, for students' test anxiety or, more specifically, schoolwork-related anxiety based on PISA 2015 data.

**Methods:** Prediction of the three machine learning methods were compared for variable selection of 188 indicators of schoolwork-related anxiety. Data were based on US students ( $N = 5593$ ) who participated in PISA 2015. With the exception of FS, LASSO, Enet, and random forest were iterated 100 times to consider the bias resulting from data-splitting to determine the selection or non-selection of each predictor. This resulted in the reporting of number of selected variables into the following five count categories: 1 or more, 25 or more, 50 or more, 75 or more, and all 100 iterations.

**Results:** LASSO and Enet both outperformed random forest but did not differ from one another in terms of prediction performance in 100 iterations of modeling. Correspondingly, LASSO was compared to FS in which, of the 188 predictors, 27 were identified as key indicators of schoolwork-related anxiety across 100 iterations, and 26 variables were also statistically significant with FS regression. Aligned with previous research, key indicators included personal, situational, and mathematics and reading achievement. Further, LASSO identified 28 variables (14.89%) statistically unrelated to schoolwork-related anxiety, which included indicators aligned to students' academic- and non-academic behaviors.

**Conclusions:** LASSO and Enet outperformed random forest and yielded comparable results in which determinants of schoolwork-related anxiety included personal and environmental factors, including achievement goals, sense of belonging, and confidence to explain scientific phenomenon. LASSO and FS also identified similar predictor variables related, as well as unrelated, to schoolwork-related anxiety. Aligned with previous research, females reported higher schoolwork-related anxiety than males. Mathematics achievement was negatively related to anxiety, whereas reading performance was positively associated with anxiety. This study also bears significance as one of the first penalized regression studies to incorporate sampling weights and reflect the complex sampling schemes of large-scale educational assessment data.

**Keywords:** Machine learning, Test anxiety, Penalized regression, Least absolute selection and shrinkage operator, Elastic net, Random forest, PISA

## Introduction

Large-scale educational assessments offer rich opportunities to investigate factors associated with a range of student outcomes. For example, the Programme for International Student Assessment (PISA) is an international assessment that gathers data from students, educators, and parents to facilitate research on the academic performance of 15-year-old students across 80 countries (Organisation for Economic Co-operation and Development [OECD], n.d.a. <https://www.oecd.org/pisa/>). A notable attribute of large-scale educational studies is the voluminous amount of data accessible to researchers seeking to select key predictors to include in statistical models of students' educational outcomes. In response, researchers have begun to apply machine learning (ML) procedures to address issues associated with big or large-scale data, including data collection and estimation (e.g., Buskirk et al., 2018; Chew et al., 2018; Eck et al., 2018; Hill et al., 2019), as well as predictive modeling (e.g., Kaur et al., 2015; Miković et al., 2019), respectively (Yoo & Rho, 2021). The application of ML procedures to large-scale educational assessment data offers much potential to advance theory and research. For example, test anxiety is a key construct in education and negatively associated with a range of student learning outcomes (e.g., academic achievement; Cizek & Burg, 2006; Chapell et al., 2005; von der Embse & Hasson, 2012; von der Embse, 2018). Within this study, penalized and non-penalized regression procedures are used for predictor variable selection of key predictors of students' schoolwork-related anxiety using US student data from PISA 2015.

Penalized, or regularized, regression is an approach to predictor variable selection of a response variable (Hastie et al., 2009). Specifically, penalized regression seeks to identify key variables to predict a response variable such as academic achievement. For instance, Least Absolute Shrinkage and Selection Operator (LASSO), a penalized regression procedure, accomplishes this by imposing a penalty parameter on the model fit function that results in variables with near-zero coefficient estimates to drop out of the model while retaining those with non-trivial effects. Penalized regression has begun to emerge as an approach to predictive modeling in diverse fields, including bioinformatics (e.g., Liu et al., 2018; Zeng et al., 2020) and engineering (e.g., Nabian & Meidani, 2020; Zhang et al., 2017). With regard to large-scale educational assessment data, elastic net (Yoo, 2018) and group mnet (Yoo & Rho, 2020) have been employed in analyzing data from the Trends in International Mathematics and Science Study (National Center for Education Statistics, n.d. <https://nces.ed.gov/timss/>) and Teaching and Learning International

Survey (OECD, n.d.b. <https://www.oecd.org/education/talis/>) studies, respectively. To-date, no studies have applied penalized regression procedures for predictor variable selection of students' test anxiety using large-scale educational assessment data.

The present study applies penalized and non-penalized regression procedures for predictor variable selection of students' test anxiety, based on data obtained from US students who participated in PISA 2015. Notably, within PISA 2015, test anxiety is operationalized in the items used to measure students' schoolwork-related anxiety, conceptualized as, "[T]he anxiety related to school tasks and tests, along with the pressure to get higher marks and the concern about receiving poor grades" (OECD, 2017, p. 84). The penalized procedures used in this study include LASSO (Tibshirani, 1996) and elastic net (hereinafter referred to as Enet; Zou & Hastie, 2005), respectively. These procedures were selected because they impose specific model constraints for predictor variable selection and to demonstrate the comparability of findings as a resource for researchers interested in these approaches. For comparative purposes, random forest (a nonlinear model utilizing decision trees) and multiple linear regression using forward stepwise (FS) were also used for predictive performance and predictor variable selection, respectively. Empirical findings serve to contribute to the limited literature on determinants of students' test anxiety (e.g., Lohbeck et al., 2016; Putwain et al., 2010) and, secondly, demonstrate the utility of alternative approaches to address substantive research questions and compare their performances with a traditional regression approach (FS) within large-scale educational assessments.

The subsequent section reviews research on students' test anxiety to illustrate its consideration as an important educational construct. This is followed by a presentation of penalized regression that addresses their key tenets and model types to show their ability to contribute to our understanding of large-scale educational assessment data.

### Test anxiety

Test anxiety is an important educational construct that has received considerable attention in the research literature over the past 70 years (e.g., Hembree, 1988; Mandler & Sarason, 1952; Spielberger et al., 1976; Zeidner, 2014). Broadly, test anxiety is, "a specific type of anxiety that can be experienced before, during, and after an examination or other test situations" (Roos et al., 2021, p. 582). It differs from anxiety disorder in that it is situation-specific, whereas the former is a general personal attribute present across contexts and situations (Cizek & Burg, 2006). As a situation-specific form of anxiety, it includes a range of interrelated reactions, including cognition, physiological, and behavioral, that students may have when they sense potential failure in an evaluative situation (e.g., high-stakes testing; Zeidner, 2014). Among school-aged children, it is estimated that the prevalence rate of test anxiety ranges between 10%-40% (Segool et al., 2013), with higher levels reported among females than males (Putwain & Daly, 2014; Seipp & Schwarzer, 1996; von der Embse et al., 2018). Among college students, the prevalence rate has been reported up to 50% among medical students (e.g., Bischofsberger et al., 2021; Tsegay et al., 2019).

Conceptually, test anxiety is a complex, multidimensional construct relevant to evaluative situations in which a student is judged according to their test performance. Specifically, test anxiety embodies the dimensions of worry and emotionality

(Spielberg & Vagg, 1995). Worry comprises the cognitive processes associated with an individuals' response to being evaluated, such as comparing one's performance to peers. The emotionality dimension includes the physiological processes associated with anxiety arousal (e.g., heart rate; Spielberg & Vagg, 1995). As a type of social-evaluation anxiety (e.g., social interactions) it encompasses affective-physiological, behavioral, and cognitive dimensions (Zeidner & Mathews, 2005). Research on the dimensions of test anxiety has resulted in the development of a range of multicomponent self-report instruments to operationalize test anxiety (e.g., Cassady & Johnson, 2002; Friedman & Bendas-Jacob, 1997; Lowe et al., 2008; Lowe, 2014; Spielberger, 1980), including their measurement invariance across groups (e.g., gender; Everson et al., 1991; Lowe, 2015).

Beyond its prevalence and measurement, research on test anxiety has focused on its determinants and correlates with cognitive and academic assessment performance. Zeidner (2014) classifies determinants of test anxiety into distal and proximal factors, with the former encompassing biological (e.g., genetics) and environmental (e.g., family) factors and the latter addressing situation-specific factors (e.g., teacher behaviors) directly associated with inducing student anxiety. Specifically, based in diverse theoretical frameworks (e.g., attachment study; Fraley, 2002), parental factors (e.g., expectations) are a source of student anxiety. Correspondingly, as a situational influence, the classroom environment (e.g., teacher regard of students) is an important factor in students' self-beliefs, including competency beliefs and motivation. For instance, classrooms in which teachers engage in normative comparisons of student performance may enhance anxiety as students may not perceive themselves as competent as other peers (Zeidner, 2014). Furthermore, students' personal beliefs (e.g., mastery—avoidance goals) have been found to be significant predictors of test anxiety (Putwain et al., 2010). For instance, lower competency beliefs have been found associated with higher test anxiety (Zeidner & Schelye, 1999). Correspondingly, increased test anxiety have been associated with higher likelihood of school dropout (Cizek & Burg, 2006), lower academic achievement and motivation (von der Embse et al., 2018), and poorer cognitive test performance (Hembree, 1988; von der Embse et al., 2018). Outcomes associated with heightened test anxiety have resulted in the development and evaluation of interventions to treat it (e.g., Putwain et al., 2014; Rose & Lomas, 2020), including targeted interventions to reduce the gender achievement gap in undergraduate STEM courses (e.g., Harris et al., 2019).

In large, research on test anxiety has focused on its prevalence among school- and college-aged samples, correlation to test performance, instrument development, and evaluation of intervention programs. Notably, the knowledge base on the ways in which environmental, situational, and personal factors influence test anxiety is a manifestation of small-scale studies focused on a few specific variables situated in theoretical models. As a result, less attention has been directed at identifying factors indicative of test anxiety to guide policy and practice decision-making. Large-scale educational assessment data, such as PISA, typically include data across a diverse range of environmental, situational, and personal factors and, thus, offer a unique opportunity to identify key determinants of text anxiety.

### Penalized regression

Penalized regression embodies a class of flexible approaches to predictor variable selection that can overcome the limitations associated with the more familiar multiple regression techniques (e.g., forward or backward stepwise regression). In the context of social science large-scale data analysis, penalized regression has several key advantages. First, penalized regression assumes linearity, and linear models are easier to interpret than nonlinear models (Yoo & Rho, 2021; Yoo et al., 2022). Although linear methods are known to produce less predictive models, in a recent study to analyze large-scale survey data, penalized regression outperformed random forest (RF), a nonlinear ML-based approach, in terms of prediction (Yoo & Rho, 2021). Second, penalized regression rests on the sparsity assumption that the true model consists of a small subset of predictors (Hastie et al., 2015). The sparsity assumption is most likely to be met using data obtained in large-scale studies than a small number of pre-selected variables typically found in small-scale studies. Third, by considering hundreds or thousands of variables that are associated with large-scale educational assessments in one statistical model, new variables or relationships can be explored and identified, which may not be feasible with traditional unpenalized regression methods (e.g., structural equation modeling, hierarchical linear modeling). Specifically, variable selection in unpenalized regression models such as ordinary least squares (OLS) regression is typically limited to a cohesive set of variables selected based on considerations related to existing theoretical frameworks and practicality (e.g., time, resources). Fourth and relatedly, penalized regression can handle high-dimensional data (i.e., large number of variables to observations) without convergence problems. Even in the social science large-scale data analysis, originally tall data (i.e., more observations than variables) often becomes high-dimensional due to the data split requirement for model validation and comparison in ML.

The basis of penalized regression, including LASSO, is the utilization of bias-variance trade-offs. That is, by introducing a small amount of bias in the regression model, variance is reduced, which ultimately leads to reduction in the prediction errors. Research has reported that LASSO combined with cross-validation is capable of selecting all true variables associated with the response variable (Bühlmann & Mandozzi, 2013; Fan & Lv, 2007). Because LASSO does not handle multicollinearity, Zou and Hastie (2005) proposed Enet, which is a combination of LASSO and ridge regression. Ridge handles multicollinearity problems by adding a penalty to the diagonal elements of a singular moment matrix (Hoerl & Kennard, 1970). As a result, Enet selects variables due to the LASSO component and handles multicollinearity due to the ridge component.

LASSO and Enet are explained using the subsequent equations. Consider a typical linear regression model,  $Y = X\beta + \epsilon$ . With  $P$  predictors and  $N$  observations,  $Y$  and  $\epsilon$  represent  $N$  dimensional vectors of a continuous response variable and error, respectively.  $X$  is an  $N$  by  $P$  matrix and  $\beta$  is a  $P$  dimensional vector of coefficients. When  $Y^T = (y_1, \dots, y_N)$ ,  $y_i$  is expressed in Equation 1.  $Y^T = (y_1, \dots, y_N)$ ,

$$y_i = \beta_0 + \sum_{j=1}^P x_{ij}\beta_j + \epsilon_i \quad (1)$$

Next, Eqs. 2 and 3 report the objective functions of LASSO and Enet for a Gaussian family, respectively. As shown, the same first term in the right-hand side of the equations is the loss function of least squares. The second term in the right-hand side of the equations is the penalty term. The penalized parameter of lambda ( $\lambda$ ) serves to control the amount of penalty. Thus, larger  $\lambda$  values result in smaller coefficients, which reduces near-zero regression coefficients to zero. Contrary, the closer  $\lambda$  gets to zero the less penalty is imposed on the coefficients. The  $\lambda$  value of 0 returns the equation to OLS. Equation 3 shows Enet includes the  $\alpha$  parameter, which bridges ridge and LASSO. The value of  $\alpha$  ranges between 0 and 1;  $\alpha$  of 0 and 1 reverts back to ridge and LASSO, respectively. Typically,  $\alpha$  is set at 0.50 (Hastie & Qian, 2016).

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^P \beta_j^2 \right\} \quad (2)$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^P (\alpha |\beta_j| + (1 - \alpha) \beta_j^2) \right\} \quad (3)$$

In the field of ML, data are divided into training and test data for model building and comparison, respectively. The tuning parameter,  $\lambda$  in the case of penalized regression, is tuned or cross-validated with the training data. The resulting optimal value of the tuning parameter is applied to test or new data, and prediction errors are calculated and compared across techniques (e.g., LASSO, random forest). Additional model details are provided in the subsequent Methods section.

### Study purpose

To-date, empirical research has established the association between text anxiety and various educational outcomes (e.g., test performance). However, less work has sought to identify predictors of test anxiety based on data obtained within the context of a large-scale study. PISA is a large-scale international assessment of 15-year-old students' academic ability (e.g., mathematics, science). Data gathered within PISA includes information on student beliefs, as well as home and school factors, to facilitate research on students' academic achievement. Because PISA gathers data on hundreds of educationally-relevant variables, it offers rich opportunities to apply regularized regression techniques to identify key classroom- and home-related indicators of students' anxiety related to their learning. This study sought to examine two research questions. The first question focused on the predictive performance of LASSO, Enet, and random forest (RF) procedures. The second question examined predictor exploration of LASSO and Enet compared to the traditional multiple regression model using FS. Notably, PISA does not measure test anxiety directly and, instead, assesses schoolwork-related anxiety, based on students' responses to five items (detailed in the subsequent *Response Variable and Data Source* section). Of the five schoolwork-related anxiety items, three items specifically address students' test anxiety. As such, a consideration associated with large-scale educational assessments is the restricted use of comprehensive assessments to measure



multidimensional, complex theoretical constructs due to a range of factors (e.g., time, resources). Previous research has supported the unidimensional structure of the PISA 2015 schoolwork-related anxiety measure and its varying degrees of measurement invariance across country by gender groups (Immekus, 2021). Study findings provide a basis to understanding factors that may be indicative of students' test-related anxiety to point to areas of future research. This includes the utility of large-scale educational assessment studies to substantiate findings from smaller scale, or targeted, research.

## Data and materials

### Response variable and data source

Data were based on the 5712 United States (US) students who participated in PISA 2015. PISA operationalizes schoolwork-related anxiety using five selected-response items. Item content is, Item 1 (ST118Q01NA): "I often worry that it will be difficult for me taking a test;" Item 2 (ST118Q02NA): "I worry that I will get poor <grades> at school;" Item 3 (ST118Q03NA): "Even if I am well-prepared for a test I feel very anxious;" Item 4 (ST118Q04NA): "I get very tense when I study for a test;" and, Item 5 (ST118Q05NA): "I get nervous when I don't know how to solve a task at school." Responses are reported on a four-point scale (1 = *Strongly Agree*; 2 = *Disagree*; 3 = *Agree*; 4 = *Strongly Agree*). Score reliability, based on Cronbach's coefficient alpha, was 0.83. From the five items, PISA provides a derived variable to indicate students' schoolwork-related anxiety, ANXTEST, which served as the response variable of the study. Seventy four percent of the sample was in grade 10 and 5593 (97.92%) completed the schoolwork-related anxiety items. The PISA 2015 data is available at <https://www.oecd.org/pisa/data/2015database/>.

### Data cleaning

Starting with 948 variables of PISA student data (928 from student questionnaire data and 20 plausible values of CPS and FLT from collaborative problem solving and financial literacy data), variables irrelevant to analyses were removed (e.g., IDs, administration variables). Of note, the final trimmed student weight (W\_FSTUWT) was the only saved weight, which was used as the sampling weight in penalized regression to account for the complex PISA data structure. Next, among 120 plausible values (PVs) of 12 subjects (10 PVs for each subject), the first PVs of each subject were retained). Further, PISA provides 54 derived variables. We retained individual items and deleted all the derived variables except ANXTEST, as it served as the response variable. Correspondingly, the five items used to derive ANXTEST (ST118Q01NA to ST118Q05NA) were removed from subsequent analysis.

PISA also provides some variables conveying similar information. For instance, variables ST006 and ST008 gather information on parents' education at different ISCED levels, whereas ST005Q1 and ST007Q1 ask the highest completed ISCED levels of parents. Intriguingly, the results were not consistent. Due to some cases of logical errors with ST006 and ST008, we included ST005Q1 and ST007Q1 in the analysis. Similarly, we deleted ST065Class, and kept ST063 on attended science courses. Last, ST003D03 on students' date of birth was merged to ST003D02T.

The measurement of PISA variables guided decisions related to their inclusion in the predictive models. Specifically, ordinal variables measured in Likert-scales were treated

as continuous, whereas binary categorical items were dummy-coded. The following multiple-categorical variables were excluded from the analysis: OCOD1 to OCOD3, ST065Class, and ST125Q01. Specifically, the occupational variables of OCOD1 to OCOD3 consisted of too many categories with extremely low responses, and collapsing of categories lacked any theoretical justification. Both ST065Class (students' science class) and ST125Q01 (ISCED levels) included the response options of 'I did not attend' and 'I do not remember,' and, thus, cannot be treated as ordinal. Last, the glmnet package in R does not handle multiple-category variables.

### Missing data imputation

Missing data for study variable were inspected prior to conducting missing data imputation. Specifically, the five variables measuring hours spent in additional learning (ST071Q01 to ST071Q05) reported the highest missing rates (range 10.98–28.25%), while the percent of missing data for the other variables ranged from 0.07 to 6.92% ( $M = 2.42$ ,  $SD = 1.64$ ). Consequently, we excluded these five variables from subsequent analysis, as the amount of missing data may impair the imputation process. Correspondingly, imputing missing values appears logically invalid for certain PISA variables. For instance, ST064 on science classes is valid only if the students attended at least one course in the school year (ST063). Students who did not attend any science courses were marked as 'valid skip' (missing) in ST064. If such variable is imputed, students who did not attend any science courses were to answer questions on science courses as if they attended them. Thus, the following variables were removed from imputation and subsequent analyses: ST064, ST098, ST100, ST103, ST104, and ST107.

As results, the final dataset consisted of 1 response variable (ANXTEST) and 188 predictors for 5593 students. Of the 188 predictor candidates, only 16 predictors (8.51%) had complete data, whereas the remaining 172 (91.49%) had missing rates ranging between 0.07% and 6.92%. Listwise deletion retained approximately 42.41% of the observations (2372 observations). We employed the  $k$ -nearest neighbors ( $k$ -NN) algorithm to handle the missing data of PISA, following Yoo and Rho (2021). After  $k$  closest observations to a certain data point are identified, the average value of the  $k$  nearest observations is replaced with the missing data point. The square root of the number of complete observations, or 49, served as the value of  $k$  (Beretta & Santaniello, 2016). Gower distance was also used in distance calculation in order to take the mixed-format data into account (Gower, 1971).

## Methods

### cross-validation and selection counts in penalized regression

In penalized regression,  $\lambda$  determines the amount of regularization and, thus, model complexity.  $K$ -fold cross-validation (CV) is a common approach to tune the penalty parameter. After training data are partitioned into  $K$  sets of equal size, the model is fitted with the training set excluding the  $k$ -th fold, and the fitted values are obtained for the  $k$ -th fold. This is repeated for every  $k$ -th fold ( $k = 1, 2, \dots, K$ ), and each fold's CV error is calculated. The average of the  $K$  folds' CV errors serves as the overall CV error (Eq. 4). The number of folds depends on sample size (Friedman et al., 2021), and 10 folds are frequently used in ML (e.g., Wu & Lange, 2008; Zhu & Hastie, 2004) and are the default of



the glmnet package (Friedman et al., 2021) in the R program (R Core Team, 2021) used in this study.

$$CV(\hat{f}, \lambda) = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{f}^{-k(i)}(x_i, \lambda)) \quad (4)$$

The following steps were used to tune the penalty parameter,  $\lambda$ . First, the data were divided into training and test data (ratio of 7:3). Second, a tenfold CV was executed with the training data to build prediction models with LASSO and Enet, respectively. The  $\lambda$  value to minimize the CV error was identified as a result. Third, using test data from Step 1, LASSO and Enet models built in Step 2 were evaluated in terms of prediction error. Specifically, root mean square error (RMSE) was calculated as the square root of the mean of squared difference of the response variable and its predicted value. RMSEs obtained with CV and test data served as the CV error and prediction error, respectively.

These three steps were iterated 100 times to consider the bias resulting from data-splitting (Meinshausen & Bühlmann, 2010; Shevade & Keerthi, 2003), which is the same as the shuffled  $k$ -fold CV in deep learning (Herent et al., 2019; Inoue et al., 2019). The selection or non-selection of each predictor in the 100 iterations was counted, which is recommended as necessary when variable selection is of interest (Yoo & Rho, 2021). This resulted in the reporting of number of selected variables into the following five categories: 1 or more, 25 or more, 50 or more, 75 or more, and all 100 iterations, respectively.

#### **Random forest and cross-validation**

RF is a nonparametric technique that yields nonlinear models by utilizing decision trees as a base learner, and is known to have strength in prediction. Specifically, RF creates bootstrapped samples, fits decision tree models on the bootstrapped samples, and combines the results of each decision tree as an ensemble method. While RF also has tuning parameters, researchers typically adopt what Breiman suggested, which are default settings in various statistical packages, including R. Specifically for a continuous response variable, Breiman (2001) suggested the number of predictors divided by 3 as the number of predictor candidates considered at each split. In a recent study, empirical findings found that efforts to tune RF parameters did not outperform Breiman's suggestion (Yoo et al., 2022). In this study, we employed the default values of the randomForest library in R (Liaw, 2022). Specifically, there were 500 trees with 62 randomly sampled variables considered as predictor candidates at each split.

We used the following steps for RF in this study. First, the same sets of training and test data (ratio of 7:3) as used in penalized regression were utilized for comparison purposes. Second, RF models were fitted with the training data. Third, the models from Step 2 were applied to the test data from Step 1. Correspondingly, the RMSE was calculated with the test data, which served as the prediction error of RF. These steps were then repeated 100 times.

#### **Forward stepwise (FS) regression**

Among classical statistical methods of variable selection including forward stepwise (FS) selection and backward stepwise selection, best subset selection is known to perform

**Table 1** Descriptive statistics of RMSE for ML

Method	Min	Q1	Median	Mean	Q3	Max	SD
LASSO	0.825	0.857	0.867	0.866	0.875	0.898	0.014
Enet	0.825	0.857	0.867	0.866	0.876	0.898	0.014
RF	0.837	0.871	0.882	0.881	0.892	0.914	0.014

RF, random forest; Q1, quartile 1; Q3, quartile 3

**Table 2** Selection counts of LASSO and Enet

Method	$\geq 1$	$\geq 25$	$\geq 50$	$\geq 75$	$= 100$
LASSO	187	149	102	79	27
Enet	187	155	103	80	27

the best, but only if the number of variables is 40 or fewer due to computational burden (Hastie et al., 2009). Bertsimas et al. (2016) provided a solution to this problem by utilizing a special optimization algorithm. Extending their work to high and low signal to noise ratio (SNR) problems, Hastie et al. (2020) showed that FS was comparable to best subset selection.

Starting with a null model, FS sequentially adds variables depending on the variable's contribution to statistical significance. Specifically, FS uses a greedy algorithm and, thus, once a variable is added to the model, it remains in the model. As FS is readily implemented with standard statistical software packages without special optimization algorithms, FS was selected among classical methods to compare with penalized regression. In this study, the step function of the stats package in R was employed with Akaike Information Criterion (AIC) as the criterion of variable selection (Step function, 2022). With linear models, this function is equivalent to the stepAIC of MASS library, which is intended for wider purposes (Venables & Ripley, 2002). FS was executed with the whole data set, and statistical significance of the variables was examined.

## Results

Table 1 reports the descriptive statistics of the root mean squared errors (RMSE) for LASSO and Enet and indicates that they were the same across procedures. Lower RMSE indicates better performance. As shown, the average and median RMSEs were 0.866 ( $SD=0.014$ ) for each approach, indicating the stability of the models across the 100 iterations, as well as producing no predictive differences. Comparatively, while the average RMSE of RF was slightly higher (Mean = 0.881,  $SD=0.014$ ), the 95% confidence intervals of the procedures overlapped. As RF did not outperform penalized regression, we opted for penalized regression in the subsequent analysis.

Table 2 reports the selection count regarding the number of times a particular variable was selected as an important predictor. Specifically, 187 out of 188 variables were selected as key predictors of schoolwork-related anxiety at least once across the 100 iterations. Correspondingly, 27 variables (14.36%) were selected as key predictors in all 100 iterations for both procedures, whereas 79 variables (42.02%) were selected with LASSO in 75 or more iterations. Furthermore, compared to LASSO, Enet tended to select

**Table 3** Descriptive statistics of variable selection in one run of LASSO and Enet

Method	Min	Q1	Median	Mean	Q3	Max	SD
LASSO	80.00	103.00	109.00	109.90	118.00	155.00	13.44
Enet	77.00	103.00	115.50	111.90	119.20	154.00	14.28

Q1, quartile 1; Q3, quartile 3

**Table 4** Numbers of selected variables of LASSO and significant variables of FS

	FS (yes)	FS (no)	
LASSO (yes)	51	28	79
LASSO (no)	5	104	109
	56	132	188

slightly more variables, but this difference subsided after applying selection counts. Specifically, LASSO selected 102 and 79 variables at least 50% and 75% of the time, whereas Enet selected 103 and 80 variables, respectively.

Table 3 presents the descriptive statistics of the number of variables selected in one run of LASSO and Enet. On average, LASSO selected 109.90 ( $SD=13.44$ ) variables, slightly less than the average of 111.90 ( $SD=14.28$ ) using Enet.

As LASSO and Enet reported almost identical results in terms of prediction error and variable selection, LASSO was selected for comparison with FS. Table 4 reports the cross-tabulation for number of variables selected using LASSO and FS. Specifically, variables selected 75 or more times by LASSO were compared with statistically significant variables of FS ( $p<0.05$ ). While the coefficient signs of variables selected 75 or more were consistent, some of the variables selected fewer than 75 times showed mixed coefficient signs. This is also consistent with the suggestion of Yoo and Rho (2021), which recommends the selection counts of 75% or more with LASSO or Enet.

Penalized regression is not scale-invariant, and standardization in the optimization algorithms is required. The R package glmnet has an option to standardize variables in optimization and put them back to the original scale, which we adopted. All the LASSO coefficients that are presented hereafter are unstandardized regression coefficients. Table 5 summarizes the 51 variables which were selected 75 or more times in LASSO and also were statistically significant with FS ( $p<0.05$ ). Specifically, the means and SDs (standard deviations) of LASSO coefficients are reported as well as the FS coefficients and the p-values. Of note, the means and SDs of LASSO coefficients were presented to provide information on the central tendency and variability of them across multiple iterations, and statistical significance should not be inferred. For instance, 0.28 and 0.02 of ST004D01T are the mean and SD of the 100 LASSO coefficients of ST004D01T, respectively. Additional techniques such as post-selection inference (Lee et al., 2016) are necessary for further significance testing.

Among the 51 variables, 26 variables were selected across all 100 iterations. As shown, regression coefficients were approximately equal across procedures with LASSO values slightly lower overall when values did differ, which can be attributed to the shrinking of LASSO coefficients. Consistent with previous research, gender

**Table 5** Statistically significant variables with FS and selected  $\geq 75$  with LASSO

Variable	Counts	LASSO mean	LASSO SD	FS coefficient	FS p-value
ST004D01T	100	0.28	0.02	0.27	0.00
ST011Q11TA	100	− 0.04	0.02	− 0.07	0.01
ST013Q01TA	100	0.02	0.01	0.05	0.00
ST119Q01NA	100	0.08	0.02	0.10	0.00
ST119Q02NA	100	0.08	0.02	0.11	0.00
ST121Q01NA	100	0.04	0.02	0.07	0.00
ST121Q03NA	100	0.08	0.02	0.11	0.00
ST082Q02NA	100	− 0.10	0.02	− 0.13	0.00
ST082Q13NA	100	0.07	0.01	0.07	0.00
ST034Q01TA	100	− 0.07	0.01	− 0.08	0.00
ST034Q04TA	100	− 0.05	0.02	− 0.05	0.01
ST034Q06TA	100	− 0.04	0.01	− 0.05	0.01
ST039Q01NA	100	0.04	0.01	0.05	0.00
ST039Q03NA	100	0.06	0.01	0.07	0.00
ST031Q01NA	100	− 0.01	0.00	− 0.01	0.03
ST032Q02NA	100	− 0.01	0.00	− 0.02	0.00
ST063Q01NB	100	− 0.06	0.02	− 0.08	0.01
ST095Q13NA	100	0.03	0.01	0.04	0.00
ST129Q02TA	100	0.04	0.01	0.05	0.01
ST129Q08TA	100	0.05	0.01	0.07	0.00
ST131Q08NA	100	0.03	0.01	0.07	0.00
ST076Q02NA	100	0.06	0.02	0.08	0.00
ST078Q02NA	100	0.20	0.03	0.22	0.00
ST078Q11NA	100	− 0.08	0.02	− 0.11	0.00
PV1MATH	100	− 0.00	0.00	− 0.00	0.00
ST016Q01NA	100	− 0.07	0.00	− 0.07	0.00
ST012Q05NA	99	0.04	0.02	0.07	0.00
ST022Q01TA	99	0.05	0.03	0.11	0.00
ST062Q01TA	99	0.04	0.02	0.06	0.00
ST129Q01TA	99	0.03	0.01	0.04	0.03
ST076Q04NA	99	− 0.05	0.02	− 0.08	0.00
ST078Q08NA	99	0.11	0.04	0.16	0.00
ST078Q09NA	99	0.05	0.02	0.06	0.01
PV1READ	98	0.00	0.00	0.00	0.00
ST123Q01NA	97	0.03	0.02	0.06	0.00
ST082Q12NA	97	0.04	0.02	0.05	0.02
ST094Q01NA	96	− 0.03	0.01	− 0.07	0.00
ST038Q04NA	96	0.03	0.01	0.04	0.05
ST001D01T	94	0.04	0.03	0.08	0.00
ST129Q06TA	94	0.02	0.01	0.04	0.03
ST078Q07NA	94	− 0.04	0.03	− 0.09	0.01
ST038Q07NA	94	0.04	0.02	0.09	0.00
ST011Q12TA	93	− 0.04	0.03	− 0.07	0.04
ST082Q03NA	91	0.03	0.02	0.05	0.04
ST012Q03TA	89	0.02	0.01	0.04	0.01
ST034Q03TA	89	0.02	0.01	0.03	0.05
PV1FLIT	88	0.00	0.00	0.00	0.01
PV1SCED	82	0.00	0.00	0.00	0.00
PV1SKCO	82	0.00	0.00	0.00	0.01
ST146Q02TA	79	0.02	0.01	0.05	0.02
ST129Q07TA	76	− 0.02	0.01	− 0.05	0.01

Unstandardized coefficients reported. Statistical significance testing should not be conducted with the means and SDs of LASSO

$p < 0.05$

differences were reported across all 100 iterations in which females reported heightened schoolwork-related anxiety compared to males. As reported, important variables spanned across such dimensions as achievement goals (e.g., ST119Q01, ST119Q02), sense of belonging (e.g., ST034Q01, ST034Q04), and science knowledge (ST131Q08, ST129Q08), as well as availability of home resources (e.g., books in the home). Furthermore, home resources including the number of books in the home (ST013Q01) and number of cell phones with internet access (ST012Q05) were positively associated with higher schoolwork-related anxiety. Notably, indicators associated with students' achievement goals were positively associated with reported anxiety. In particular, increased anxiety was associated with wanting top grades in most or all courses (ST119Q01), selecting from among the best opportunities available when they graduate (ST119Q02), and wanting top grades at school and continue working on tasks until perfect (ST121Q03), respectively.

Furthermore, increased study behaviors, such as engaging in homework or studying before (ST076Q02) or after school (ST078Q02), were also associated with increased anxiety. Notably, students who reported feeling like an outsider (ST034Q01) or awkward and out of place in their school (ST034Q04) reported less anxiety. However, higher anxiety was associated with students who reported that their teachers called on them less often than other students (ST039Q01) and gave students the impression that they think they are less smart than they really are (ST039Q03). Interestingly, several items related to students' science knowledge were also associated with higher anxiety. This included the item asking students to explain why earthquakes occur more frequently in some areas than others (ST129Q02) and identify the better of two explanations for the formation of acid rain (ST29Q08). Unsurprisingly, engagement in vigorous physical activities for at least 20 min per day (ST032Q02) and exercising or practicing a sport after school (ST078Q11) were both negatively associated with student anxiety. Likewise, students who reported meeting friends or talking to friends on the phone (ST078Q07) also reported lower anxiety, including those who reported being satisfied with their life as a whole (ST016Q01). Notably, increased mathematics achievement was negatively associated with reported anxiety (coefficient =  $-0.005$ ), whereas reading achievement was positively associated with anxiety (coefficient =  $0.002$ ), respectively.

Correspondingly, LASSO and FS identified 28 variables (14.89%) selected 75 times or more but statistically unrelated with schoolwork-related anxiety. Table 6 indicates these variables related to students' academic- and non-academic behaviors (e.g., ST078Q03NA: "After leaving school did you: Watch TV\<DVD>\Video"), home resources (e.g., ST011Q01TA: "In your home: A desk to study at"), and perceptions of teacher behaviors (e.g., ST039Q05NA: "Teachers ridiculed me in front of others."). Furthermore, 16 of the 28 variables (57.14%) were not included in the FS modeling (e.g., ST123Q02NA: "My parents support my educational efforts and achievement.").

Table 7 reports the 5 variables (2.66%) selected as related to schoolwork-related anxiety less than 75 times with LASSO and FS ( $p < 0.05$ ). With the exception of ST032Q01NA ("Moderate physical activities for a total of at least 60 min per day"), the remaining variables included Science subscale scores.

**Table 6** Statistically non-significant variables with fs and selected  $\geq 75$  with LASSO

Variable	Counts	LASSO mean	LASSO SD	FS coefficient	FS p-value
ST082Q01NA	96	0.02	0.01	0.02	0.20
ST082Q09NA	91	0.02	0.01	0.03	0.12
ST039Q05NA	78	0.01	0.01	0.03	0.11
ST076Q01NA	81	− 0.02	0.00	− 0.04	0.11
PV1CLPS	100	0.00	0.00	0.00	0.08
ST078Q03NA	78	− 0.03	0.02	− 0.06	0.07
ST063Q06NA	91	− 0.04	0.02	− 0.06	0.07
ST092Q02TA	90	− 0.02	0.01	− 0.02	0.07
ST011D17TA	93	− 0.03	0.02	− 0.05	0.06
ST063Q03NB	79	0.02	0.01	0.05	0.06
ST060Q01NA	81	0.00	0.00	0.00	0.06
ST038Q03NA	99	0.04	0.01	0.04	0.05
ST123Q02NA	95	0.03	0.02	−	−
ST019CQ01T	95	0.03	0.02	−	−
ST012Q01TA	93	0.02	0.01	−	−
ST039Q02NA	91	0.01	0.01	−	−
ST082Q08NA	90	0.02	0.02	−	−
ST034Q02TA	88	0.02	0.01	−	−
ST119Q03NA	86	0.02	0.02	−	−
ST078Q06NA	86	− 0.02	0.02	−	−
ST012Q09NA	85	0.01	0.01	−	−
ST131Q01NA	84	0.01	0.01	−	−
ST121Q02NA	81	0.01	0.01	−	−
ST146Q05TA	81	0.02	0.02	−	−
ST011Q01TA	80	0.02	0.02	−	−
ST076Q11NA	78	− 0.02	0.02	−	−
ST078Q10NA	78	− 0.02	0.02	−	−
ST095Q15NA	75	0.01	0.01	−	−

Unstandardized coefficients reported

PV1CLPS, plausible value 1 of collaborative problem solving. Statistical significance testing should not be conducted with the means and SDs of LASSO

$p > 0.05$

**Table 7** Statistically significant variables with FS and selected  $< 75$  with LASSO

Variable	Counts	LASSO mean	LASSO SD	FS coefficient	FS p-value
Moderate physical activities at least 60 min per day (ST032Q01NA)	60	0.0025	0.00	0.0118	0.0459
Explain phenomena scientifically science subscale (PV1SCEP)	25	− 0.0001	0.00	− 0.0013	0.0010
Living system subscale (PV1SSLI)	22	0	0.00	− 0.0011	0.0029
Science (PV1SCIE)	14	0	0.00	0.0012	0.0107
Physical science subscale (PV1SSPH)	9	0	0.00	0.0010	0.0078

SDs less than 0.00. PV = Plausible value. Statistical significance testing should not be conducted with the means and SDs of LASSO

$p < 0.05$



## Discussion

The aim of this study was to examine the utility of penalized regression procedures for predictor variable selection of US students' schoolwork-related anxiety based on 188 student variables in PISA 2015. Particular focus of this study was addressing the literature gap on the use of methodological advancements in ML to identify key indicators of test anxiety. Despite test anxiety being widely considered an important construct in education it has received surprisingly little consideration in the context of large-scale educational assessments. Correspondingly, there is little available information regarding key predictors of students' anxiety within educational settings that can be used to guide further research. In response, this study examined the performance of LASSO, Enet, and random forest for variable selection of indicators of students' test anxiety, operationalized by the PISA variable as schoolwork-related anxiety. Notably, this study bears significance as one of the first penalized regression studies to incorporate sampling weights and, thus, reflect the complex sampling schemes of social science large-scale data. In general, study findings were largely consistent with previous findings and shed light on previously unexamined variables as key indicators of schoolwork-related anxiety and offer directions for future research.

To begin, LASSO and Enet procedures produced similar results and outperformed random forest. Regardless of penalized regression procedures, 187 variables out of 188 were identified as a key predictor at least once in each of the 100 iterations. Aligned with previous research (Yoo & Rho, 2021), this asserts the advantage of employing multiple iterations and selection counts to identify key predictors of a response variable with the use of penalized regression. As with large-scale educational assessment data, the use of variable counts to identify significant predictors across iterations is particularly important when variable selection is of primary interest with many indicators. Furthermore, despite the additional model constraint in Enet to address multicollinearity concerns this did not result in appreciable discrepancies in empirical findings. As a result, researchers may consider the use of LASSO or Enet procedures when pursuing similar analyses with PISA 2015 data. Furthermore, the similarity of findings between penalized regression and FS to identify the weakest and strongest predictors of schoolwork-related anxiety indicates either approach may be used in instances in which research seeks to identify the variables most strongly or weakly associated with the outcome of interest (e.g., academic achievement). Additional research is needed to determine the generalizability of these findings to other PISA studies and large-scale educational assessment data.

To situate empirical findings to previous research requires consideration of the approach of large-scale educational assessment studies, such as PISA, to the operationalization of study variables. Specifically, test anxiety is a multidimensional construct typically measured using multicomponent, self-report instruments comprising conceptual subscales (e.g., physiological anxiety, test anxiety; e.g., Lowe, 2014; Reynolds et al., 2003; Spielberger, 1980). Such measures are developed according to theoretical and empirical considerations with subsequent validation research focused on substantiating score interpretation and use for designated purposes, including testing an instrument's measurement invariance across groups (e.g., gender; Everson et al., 1991; Lowe, 2015). Contrary, the scope of large-scale educational assessments generally restricts the measurement of key constructs to single- or multi-item scales to capture diverse

aspects of students' educational experiences to facilitate cross-national studies on academic achievement. Consequently, due to the administration of hundreds of questions, the operationalization of such constructs as test anxiety may not be as exhaustive as the aforementioned instruments. Specifically, PISA 2015 captures test anxiety within the item set designed to assess students' schoolwork-related anxiety. While the collective item set asks students to rate the degree to which they worry and are anxious about test taking, the items are not organized into theoretical dimensions of test anxiety. Thus, a key consideration of interpreting the meaningfulness of reported results is the measurement of study variables, particularly when findings may or may not align with previously reported results. For the schoolwork-related anxiety measure, its factor structure has been found to be unidimensional with varying degrees of measurement invariance across countries and gender (Immekus, 2021).

In large, study findings align with previous results on factors associated with students' test anxiety. Specifically, consistent with meta-analytic findings (e.g., Hembree, 1988; von der Embse et al., 2018), females reported higher schoolwork-related anxiety compared to males. Significant variables aligned with students' academic-related self-beliefs fell within such dimensions as achievement goals, sense of belonging, and confidence to explain scientific phenomenon. In particular, schoolwork-related anxiety was positively associated with students' achievement goals of wanting top grades in most or all of their courses, ability to select from the best opportunities available upon graduation, and wanting to get top grades at school and continue working on tasks until perfect and schoolwork-related anxiety, respectively. The positive association of students' future oriented competency-related outcomes are aligned with von der Embse et al.'s (2018) meta-analysis that found a positive, medium correlation ( $r=0.30$ ) with mastery-avoidance goals and a small, positive correlation with performance-approach goals ( $r=0.09$ ). Likewise, heightened anxiety was positively associated with students who reported giving up easily when confronted with a problem and often not prepared. Relatedly, students' studying behaviors were positively associated with schoolwork-related anxiety. Specifically, students who reported studying before going to school, or after leaving school, reported higher schoolwork-related anxiety. In consideration of sense of belonging, students who reported feeling like an outsider and awkward and out of place in their school reported lower schoolwork-related anxiety, suggesting that students who feel a lack of connection to one's school are less anxious about their schoolwork. Such findings tie into key environmental factors associated with the anxiety of students in academic settings (Zeidner, 2014). On the other hand, students who reported that other students left them out of things on purpose had higher anxiety. Notably, students with reported higher levels of agreement for explaining scientific phenomenon (e.g., frequency of earthquake occurrences in some areas, identify better explanation of the formation of acid rain) also reported higher schoolwork-related anxiety. The basis of this association between perceived ability to explain scientific phenomenon and anxiety points to an area of future research, particularly how it may tie into previous research. These findings point to subsequent research on the way in which students' academic self-beliefs facilitate the attainment of their achievement goals and accompanying anxiety associated with meeting their standards for success.

Test anxiety is routinely reported to be negatively related with academic performance (e.g., grades, standardized tests; Hembree, 1988; von der Embse et al., 2018). In the present study, mathematics achievement was negatively associated with students' schoolwork-related anxiety, whereas reading performance was positively associated with reported anxiety. Notably, the negative association between mathematics performance and schoolwork-related is closely aligned with previous findings (e.g., Segool et al., 2013). However, the positive association between reading test performance and schoolwork-related anxiety suggests that higher test performance is not necessarily related to lower anxiety levels. Further, the PISA measure of schoolwork-related anxiety does not explicitly assess the degree anxiety may be debilitating. Furthermore, previous research has considered the relationship between test anxiety and high-stakes assessments, whereas PISA is a low-stakes educational assessment due to the lack of consequences associated with test performance. In PISA 2015, concentration was on the content domain of science with mathematics and reading as minor assessment domains (OECD, n.d.a. <https://www.oecd.org/pisa/>). In this study, science test performance was not a significant indicator in each of the 100 iterations. However, four science-related subscale scores (e.g., physical science, living system) were found to be related to schoolwork-related anxiety 25 or less times across LASSO and FS procedures. Helmke (1988) reported the large differences in the correlation between test anxiety and academic achievement (range: 0.38 to  $-0.82$ ) indicating that heightened anxiety may be associated with improved test performance but at a certain point may become debilitating. Correspondingly, Helmke identified specific classroom factors (e.g., teacher task orientation, incentive value of achievement) that may moderate the relationship between test anxiety and academic performance. In this study, students who reported that their teachers called on them less often than other students and gave them the impression that they were less smart than they really were had higher schoolwork-related anxiety, which ties to the ways in which teachers' evaluative standards may influence student anxiety (Zeidner, 2014). Future research may further examine the relationship between test anxiety and academic performance that consider alternative moderators (e.g., classroom factors) that may account for the varying relationship between these variables.

Empirical findings shed additional light on the association between students' physical activity, social engagements, and anxiety. In particular, vigorous physical activities for at least 20 min per day and exercising or practicing a sport after school were negatively associated with anxiety. Likewise, students who reported being more satisfied with their life as a whole these days had reduced anxiety. This aligns with Andermo et al.'s (2020) meta-analysis of the positive effects of school-related physical activity interventions on children's mental health. In particular, reduced anxiety (Hedges'  $g = 0.347$ , 95% CI 0.072–0.623) and heightened well-being (Hedges'  $g = 0.877$ , 95% CI 0.356–1.398) were two of the four mental health outcomes (e.g., resilience) that benefited from physical activity. Furthermore, students' engagement in moderate physical activities at least 60 min per day (ST032Q01NA) was found to be a significant predictor of schoolwork-related anxiety in 60 iterations. In addition, students who meet friends or talked to friends on the phone, as well as indicated that they are a good listener, also had lower reported anxiety. As such, the degree to which students engage in consistent physical activity and a peer group appear to be protective factors of their

schoolwork-related anxiety. Such findings indicate the complexity of factors that are associated with the anxiety levels of students in academic settings, including areas of research on understanding the way in which students' out-of-school experiences shape their mental health in school settings.

Notwithstanding the application of ML-based techniques to large-scale educational assessment data there are several considerations regarding the implication of study findings. As addressed, test anxiety was operationalized according to the PISA 2015 schoolwork-related anxiety. Consequently, empirical findings should be interpreted in consideration of how the construct of anxiety is measured within PISA compared to smaller scale studies that use multicomponent instruments of test anxiety. Furthermore, as with FS, the regularized regression procedures used in this study are data-driven procedures designed to identify key indicators of a response variable. That is, many variables were consistently identified as key predictors of students' anxiety in which future research may consider their inclusion as potential moderators (e.g., physical activity level, classroom variables) in theoretically-based models to examine its association with various student learning outcomes (e.g., academic achievement). As such, empirical findings provide a basis for continued research on factors associated with students' schoolwork-related anxiety.

Penalized regression procedures offer a range of opportunity to examine the association of the vast amount of data collected within large-scale educational assessment studies. As provided in this study, application of these regression procedures across 100 iterations indicated the researchers should expect LASSO and Enet to yield similar findings and outperform random forest. Likewise, there was general agreement in LASSO and FS procedures regarding those indicators with the strongest and weakest association with schoolwork-related anxiety. Thus, depending on the study's intent, either method may be considered for use to identify those indicators to retain or exclude from model building. Of course, this suggestion is limited to the current data and future research is needed to further understand instances in which penalized regression and FS methods may produce divergent findings. Of note, FS is also data-driven; when we randomly deleted 10% of the data and ran FS with the remaining 90% of the data, the results were different from using the full data. LASSO is an improvement from the traditional methods such as FS in that it applies a soft threshold, compared to the hard threshold of FS, and identifies true predictor candidates, when combined with cross-validation (Bühlmann & Mandozzi, 2013; Fan & Lv, 2007). By conducting the analyses across iterations a resultant count provides a basis to disentangle the salience between predictor variables based on the number of times that are flagged as key indicators. Of the 188 predictor variables considered, 27 were identified as important indicators in each of the 100 iterations, 26 of which were also statistically significant with FS, yielding empirical evidence on indicators for consideration in future research seeking to understand factors associated with students' schoolwork-related anxiety. Future research is needed to further understand the contribution of regularized regression techniques in large-scale educational assessment to guide policy and research decision-making.

## Abbreviations

ML	Machine learning
Enet	Elastic net
RF	Random forest
FS	Forward stepwise regression

## Acknowledgements

Not applicable.

## Author contributions

JI contributed to the writing of the manuscript and interpretation of the data. TJ analyzed the data, and supported the interpretation of the results. JY designed and lead the study. All authors read and approved the final manuscript.

## Funding

No funding supported this study.

## Availability of data and materials

The datasets generated and/or analysed during the current study are available in the PISA 2015 repository, <https://www.oecd.org/pisa/data/2015database/>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All the authors consented for publication.

### Competing interests

The authors declare that they have no competing interests.

Received: 30 October 2021 Accepted: 12 December 2022

Published online: 27 December 2022

## References

- Andermo, S., Hallgren, M., Nguyen, T.-T.-D., Jonsson, S., Petersen, S., Friberg, M., Rornqvist, A., Stubbs, B., & Elinder, L. S. (2020). School-related physical activity interventions and mental health among children: A systematic review and meta-analysis. *Sports Medicine—Open*, 6, 25. <https://doi.org/10.1186/s40798-020-00254-x>
- Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: A critical evaluation. *BMC Medical Informatics and Decision Making*, 16(S3), 97–208. <https://doi.org/10.1186/s12911-016-0318-z>
- Bischofsberger, L., Burger, P. H. M., Hammer, A., Paulsen, F., Scholz, M., & Hammer, C. M. (2021). Prevalence and characteristics of test anxiety in first year anatomy students. *Annals of Anatomy*, 236, 151719. <https://doi.org/10.1016/j.aanat.2021.151719>
- Breiman, L. (2001). Random forests. *Machine Learning*, 54(1), 5–32.
- Bühlmann, P., & Mandozzi, J. (2013). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Computational Statistics*. <https://doi.org/10.1007/s00180-013-0436-3>
- Buskirk, T. D., Bear, T., & Bareham, J. (2018, October 25–27). *Machine made sampling designs: Applying machine learning methods for generating stratified sampling designs* [Paper presentation]. [Conference presentation] Big Data Meets Survey Science Conference, Barcelona, Spain.
- Cassady, J. C., & Johnson, R. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27, 270–295. <https://doi.org/10.1006/ceps.2001.1094>
- Chapell, M. S., Blanding, Z. B., Silerstein, M. E., Takahashi, M. N. B., Newman, B., Gubi, A., & McCain, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology*, 97, 268–274. <https://doi.org/10.1037/0022-0663.97.2.268>
- Chew, R., Jones, K., Unangst, J., Cajka, J., Allpress, J., Amer, S., & Krotki, K. (2018). Toward model-generated household listing in low-and middle-income countries using deep learning. *ISPRS International Journal of Geo-Information*, 7(11), 448. <https://doi.org/10.3390/ijgi7110448>
- Cizek, G., & Burg, S. S. (2006). *Addressing test anxiety in a high-stakes environment*. Sage.
- Eck, A., Buskirk, T., Fletcher, K., Stefek, P., Shao, H., Park, K., & Losch, M. (2018, October 25–27). *Machine made sampling frames: Creating sampling frames of windmills and other non-traditional sampling units using machine learning with neural networks* [Paper presentation]. [Conference presentation] Big Data Meets Survey Science Conference, Barcelona, Spain.
- Everson, H. T., Millsap, R. E., & Rodriguez, C. M. (1991). Isolating gender differences in text anxiety: A confirmatory factor analysis of the Test Anxiety Inventory. *Educational and Psychological Measurement*, 51, 243–251. <https://doi.org/10.1177/0013164491511024>
- Fan, J., & Lv, J. (2007). Sure independence screening for ultra-high dimensional feature space. *Journal of the Royal Statistical Society: Series B (methodological)*, 70(5), 849–911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay K, Simon, N., & Qian, J. (2021). *Package 'glmnet'* <https://cran.r-project.org/web/packages/glmnet/glmnet.pdf>

- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27, 857–871. <https://doi.org/10.2307/2528823>
- Harris, R. B., Grunspan, D. Z., Pelch, M. A., Fernandes, G., Ramirez, G., & Freeman, S. (2019). Can test anxiety interventions alleviate a gender gap in an undergraduate STEM course? *CBE-Life Sciences Education*, 1, 1–9. <https://doi.org/10.1187/cbe.18-05-0083>
- Hastie, T., & Qian, J. (2016). *Glmnet vignette*. [https://hastie.su.domains/Papers/Glmnet\\_Vignette.pdf](https://hastie.su.domains/Papers/Glmnet_Vignette.pdf)
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.
- Hastie, T., Tibshirani, R., & Tibshirani, R. J. (2020). Best subset, forward stepwise or LASSO? Analysis and recommendations based on extensive comparisons. *Statistical Science*, 35, 579–592. <https://doi.org/10.1214/19-STS733>
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58, 47–77.
- Herent, P., Schmauch, B., Jehanno, P., Dehaene, O., Saillard, C., Balleyguier, C., Arfi-Rouche, J., & Jegou, S. (2019). Detection and characterization of MRI breast lesions using deep learning. *Diagnostic and Interventional Imaging*, 100(4), 219–225. <https://doi.org/10.1016/j.diii.2019.02.008>
- Hill, C. A., Biemer, P., Buskirk, T., Callegaro, M., Cazar, A. L. C., Eck, A., Japac, L., Kirchner, A., Kolenikov, S., Lyberg, L., & Sturgis, P. (2019). Exploring new statistical frontiers at the intersection of survey science and big data: Convergence at “Big-Surv18”. *Survey Research Methods*, 13(1), 123–134. <https://doi.org/10.18148/srm/2019.v1i1.7467>
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Immekus, J. C. (2021). Multiple groups CFA and alignment approaches for testing measurement invariance and factor score estimation: Illustration with the schoolwork-related anxiety survey across countries and gender. *Methodology*, 17, 22–38. <https://doi.org/10.5964/meth.2281>
- Inoue, T., Vinayavekhin, P., Wang, S., Wood, D., Munawar, A., Ko, B. J., Greco, N., & Tachibana, R. (2019). *Domestic activities classification based on CNN using shuffling and mixing data augmentation* [Paper presentation]. [Conference presentation] Detection and Classification of Acoustic Scenes and Events 2019, New York, NY. [http://dcase.community/documents/workshop2019/proceedings/DCASE2019Workshop\\_Inoue\\_20.pdf](http://dcase.community/documents/workshop2019/proceedings/DCASE2019Workshop_Inoue_20.pdf)
- Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science*, 57, 500–508. <https://doi.org/10.1016/j.procs.2015.07.372>
- Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3), 907–927.
- Liaw, A. (2022). *Package ‘randomForest’*. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- Liu, J., Liang, G., Siegmund, K. D., & Lewinger, J. P. (2018). Data integration by multi-tuning parameter elastic net regression. *BMC Bioinformatics*, 19, 369. <https://doi.org/10.1186/s12859-018-2401-1>
- Lohbeck, A., Nitkowski, D., & Petermann, F. (2016). A control-value theory approach: Relationships between academic self-concept, interest, and test anxiety in elementary school children. *Child and Youth Care Forum*, 45, 887–904. <https://doi.org/10.1007/s10566-016-9362-1>
- Lowe, P. A. (2014). The test anxiety measure for adolescents (TAMA): Examination of the reliability and validity of the scores of a new multidimensional measure of test anxiety for middle and high school students. *Journal of Psychoeducational Assessment*, 32, 404–416. <https://doi.org/10.1177/0734282913520595>
- Lowe, P. A. (2015). Should test anxiety be measured differently for males and females? Examination of measurement bias across gender on measures of test anxiety for middle and high school, and college students. *Journal of Psychoeducational Assessment*, 33, 238–246. <https://doi.org/10.1177/0734282914549428>
- Lowe, P. A., Lee, S. W., Witteborg, K. M., Prichard, K. W., Luhr, M. E., Cullinan, C. M., Mildren, B. A., Raad, J. M., Cornelius, R. A., & Janik, M. (2008). The Text Anxiety Inventory for Children and Adolescents (TAICA): Examination of the properties of a new multidimensional measure of test anxiety among elementary and secondary school students. *Journal of Psychoeducational Assessment*, 26, 215–230. <https://doi.org/10.1177/0734282907303760>
- Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, 47, 166–173.
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (methodological)*, 72(4), 417–473. <https://doi.org/10.1111/j.1467-9868.2010.00740.x>
- Miković, R., Arsić, B., Gligorić, D., Gačić, M., Petrović, D., & Filipović, N. (2019). The influence of social capital on knowledge management maturity of nonprofit organizations—Predictive modelling based on a multilevel analysis. *IEEE Access*, 7, 47929–47943. <https://doi.org/10.1109/access.2019.2909812>
- Nabian, M. A., & Meidani, H. (2020). Physics-driven regularization of deep neural networks for enhanced engineering design and analysis. *Journal of Computing and Information Science in Engineering*, 20, 011006. <https://doi.org/10.1115/1.4044507>
- National Center for Education Statistics (n.d.). *Trends in International Mathematics and Science Study*. <https://nces.ed.gov/timss/>
- Organisation for Economic Co-operation and Development (2017). *PISA 2015 results (volume III): Students’ well-being*, PISA, OECD Publishing, Paris. <https://doi.org/10.1787/9789264273856-en>
- Organisation for Economic Co-operation and Development (n.d.a) *Program for International Student Assessment*. <https://www.oecd.org/pisa/>
- Organisation for Economic Co-operation and Development (n.d.b) *Teaching and Learning International Survey*. <https://www.oecd.org/education/talis/>
- Putwain, D. W., Chamberlain, S., Daly, A., & Sadreddini, S. (2014). Reducing test anxiety among school-aged adolescents: A field experimental. *Educational Psychology in Practice*, 30, 420–440. <https://doi.org/10.1080/02667363.2014.964392>
- Putwain, D. W., & Daly, A. L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies*, 40, 554–570. <https://doi.org/10.1080/03055698.2014.953914>
- Putwain, D. W., Woods, K. A., & Symes, W. (2010). Personal and situational predictors of test anxiety of students in post-compulsory education. *British Journal of Educational Psychology*, 80, 137–160. <https://doi.org/10.1348/000709909X466082>
- R Core Team (2021). *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>



- Reynolds, C. R., Richmond, B. O., & Lowe, P. A. (2003). *The adult manifest anxiety scale—College version*. Western Psychological Services.
- Roos, A.-L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review*, 33, 579–618. <https://doi.org/10.1007/s10648-020-09543-z>
- Rose, S. E., & Lomas, M. H. R. (2020). The potential of a mindfulness-based coloring intervention to reduce test anxiety in adolescents. *Mind, Brain, and Education*, 14, 335–340.
- Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened test anxiety among young children: Elementary school students' anxious responses to high-stakes testing. *Psychology in the Schools*, 50, 489–499. <https://doi.org/10.1002/pits.21689>
- Seipp, B., & Spielberger, C. (1996). Cross-cultural anxiety: A review. In C. Schwarzer & M. Zeidner (Eds.), *Stress, anxiety, and coping in academic settings* (pp. 13–68). Francke-Verlag.
- Shevade, S. K., & Keerthi, S. S. (2003). A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics (Oxford, England)*, 19(17), 2246–2253. <https://doi.org/10.1093/bioinformatics/btg308>
- Spielberger, C. D. (1980). *Test Anxiety Inventory: Preliminary Professional Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C. D., Anton, W., & Bedell, J. R. (1976). The nature and treatment of test anxiety. In M. Zuckerman & C. D. Spielberger (Eds.), *Emotions and anxiety: New concepts, methods and applications* (pp. 317–344). Lawrence Erlbaum.
- Spielberger, C. D., & Vagg, P. R. (Eds.). (1995). *Test anxiety: Theory, assessment, and treatment*. Taylor & Francis.
- Step Function. (2022). Choose a model by AIC in a stepwise algorithm. <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/step.html>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tsegay, L., Shumet, S., Damene, W., Gebreegziabhier, G., & Ayano, G. (2019). Prevalence and determinants of test anxiety among medical students in Addis Ababa Ethiopia. *BMC Medical Education*, 19, 423. <https://doi.org/10.1186/s12909-019-1859-5>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
- von der Embse, N., & Hasson, R. (2012). Test anxiety and high-stakes test performance between school settings: Implications for Educators. *Preventing School Failure*, 56, 180–187. <https://doi.org/10.1080/1045988X.2011.633285>
- von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, 227, 483–493. <https://doi.org/10.1016/j.jad.2017.11.048>
- Wu, T. T., & Lange, K. (2008). Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics*, 2(1), 224–244.
- Yoo, J. E. (2018). TIMSS 2011 Student and teacher predictors for mathematics achievement explored and identified via elastic net. *Frontiers in Psychology*, 9, 317. <https://doi.org/10.3389/fpsyg.2018.00317>
- Yoo, J. E., & Rho, M. (2020). Exploration of predictors for Korean teacher job satisfaction via a machine learning technique, Group Mnet. *Frontiers in Psychology*, 11, 441. <https://doi.org/10.3389/fpsyg.2020.00441>
- Yoo, J. E., & Rho, M. (2021). Large-scale survey data analysis with penalized regression: A Monte Carlo simulation on missing categorical predictors. *Multivariate Behavioral Research*. <https://doi.org/10.1080/00273171.2021.1891856>
- Yoo, J. E., Rho, M., & Lee, Y. (2022). Online students' learning behaviors and academic success: An analysis of LMS log data from flipped classrooms via regularization. *IEEE Access*, 10, 10740–10753. <https://doi.org/10.1109/ACCESS.2022.3144625>
- Zeidner, M. (2014). Anxiety in education. In R. Pekrun & E. A. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 265–288). Taylor & Francis.
- Zeidner, M., & Matthews, G. (2005). Evaluation anxiety: Current theory and research. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 141–163). New York: Guilford Publications.
- Zeng, C., Thomas, D. C., & Lewinger, J. P. (2020). Incorporating prior knowledge into regularized regression. *Bioinformatics*, 37, 514–521. <https://doi.org/10.1093/bioinformatics/btaa776>
- Zhang, Y., Minchin, E., Jr., & Agdas, D. (2017). Forecasting completed cost of highway construction projects using LASSO regularized regression. *Journal of Construction Engineering and Management*, 143(10), 04017071.
- Zhu, J., & Hastie, T. (2004). Classification of gene microarrays by penalized logistic regression. *Biostatistics*, 5(3), 427–443.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67, 301–320.
- Zeidner, M., & Schleyer, E. (1999). The big-fish-little pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology*, 24, 305–329. <https://doi.org/10.1006/ceps.1998.0985>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.