2021 ■ Vol. 17 ■ no. 3

Who Belongs in School? Using Statistical Learning Techniques to Identify Linear, Nonlinear and Interactive Effects

Rafael Quintana [a]

[a] School of Education and Human Sciences, University of Kansas

Abstract ■ The sense of school belonging refers to students' feelings of being accepted and connected to their particular school. School belonging has been considered an important determinant of a range of academic and socioemotional outcomes. Yet despite an extensive literature on the topic, it is not clear what factors are more strongly related to the students' sense of school belonging. Using a nationally representative dataset, we investigated the extent to which school belonging in fifth grade can be predicted by a wide range of individual and contextual-level factors using two statistical learning techniques (Lasso and MARS). The strongest predictor of school belonging across all models was students' feelings of peer social support, followed by students' feelings of loneliness at school. These results suggest that peer social relationships are a key component of students feeling of being connected to their school.

Keywords ■ school belonging, statistical learning, Lasso, MARS.

Acting Editor ■ Cheng-Ta Yang (Department of Psychology, National Cheng Kung University)

Reviewers ■ Two anonymous reviewers.

quintana@ku.edu

10.20982/tqmp.17.3.p312

Introduction

The sense of school belonging refers to students' feelings of fitting in, i.e., of being accepted and connected to their particular school (Anderman, 2003). School belonging has been considered an important determinant of a range of academic and socioemotional outcomes (Anderman & Freeman, 2004). Yet despite an extensive literature on this topic, it is not clear what factors are more important in promoting students' sense of school belonging. Causal research in this field is scarce, and correlational studies point to a wide-range of potential factors related to the student, teachers, families, peers and schools (e.g. Allen, Kern, Vella-Brodrick, Hattie, & Waters, 2018; Korpershoek, Canrinus, Fokkens-Bruinsma, & de Boer, 2019).

In a recent meta-analysis, Allen et al. (2018) report that eight general 'themes' are significantly associated with school belonging: academic motivation, parent support, teacher support, emotional stability, peer support, gender, personal characteristics and environmental features. The authors use the word 'theme' to indicate that each of these dimensions is composed of many different factors. For example, 'personal characteristics', 'environmental features' or 'emotional stability' can be operationalized in many different ways. Prior research suggests, then, that many variables related to different dimensions can potentially influence students' sense of school belonging. This is a common finding in social and behavioral research, where outcomes are typically related to many factors –they are, to use Fried and Robinaugh's (2020) expression, 'massively multifactorial'.

Once we recognize that there can be dozens or even hundreds of potentially informative factors related to our outcome of interest, two important questions arise. First, to what extent can we predict the outcome of interest using the observed factors? And second, which factors are stronger predictors of the outcome? Even if social and behavioral researchers have been mostly concerned with describing causal mechanisms, assessing predictability has important practical and scientific consequences (Hofman, Sharma, & Watts, 2017; Yarkoni & Westfall, 2017). Among other things, predictive modelling can be used to inform policy interventions (Kleinberg, Ludwig, Mullainathan, & Obermeyer, 2015), and lead to the generation of new theories, methods, measures and hypotheses (Hofman et al., 2017; Shmueli, 2010; Yarkoni & Westfall, 2017).

The explanatory or 'model-based' approach to analyzing data generally operates by assuming a parametric (typically linear) model that connect a set of inputs to a particular output (Breiman, 2001; Molina & Garip, 2019). The proposed model is supposed to replicate the data-generating process, in which case the parameter estimates obtained by fitting the model to the data may be unbiased. Even if this approach can be successful in different scenarios, it also has important disadvantages. Notably, the results depend on the particular model assumed by the researcher. The fact that researchers work with different datasets and choose one among potentially many different models is the source of well-known problems in social sciences such as 'p-hacking' or 'researcher degrees of freedom' (Gelman & Loken, 2013; Simmons, Nelson, & Simonsohn, 2011).

The increasing amount of rich and large datasets as well as developments in statistical learning have opened the door to investigate more flexible relationships than the ones implied by the linear and additive models commonly used in social and behavioral research (e.g. Molina & Garip, 2019; Varian, 2014). In addition, statistical learning techniques allow researchers to compare the predictive accuracy of different models as well as classes of models.

For example, statistical learning methods allow us to investigate whether including interaction terms improves the predictive accuracy of the model. Social and behavioral phenomena are generally considered context-dependent, in the sense that they are moderated by individual and contextual-level factors (e.g. Cartwright & Hardie, 2012; Ragin, 2009). For instance, Allen et al. (2018) report that the effect of several factors on school belonging is moderated by geographic location (the effects are generally stronger in rural and suburban areas). Yet a common problem when examining interactions is that only some interactions are considered, and interaction terms increase the risk of overfitting (e.g. Simmons et al., 2011). Statistical learning techniques allow us to search for arbitrarily complex interactions between predictors while preventing over-fitting the model by assessing the out-of-sample predictive performance (e.g. Hastie, Tibshirani, & Friedman, 2009; Mullainathan & Spiess, 2017a).

The goals of this paper are twofold. First, we investigate the extent to which students' sense of school belonging can be predicted by a wide range of individual and contextual-level variables using two statistical learning models: the least absolute shrinkage and selection operator (Lasso), and multivariate adaptive regression splines (MARS). Second, we use these models to identify which individual or contextual-level factors are more predictive of school belonging.

We use a nationally representative sample (ECLS-K: 2011) of 11,434 children and focus on students' sense of school belonging in fifth grade. Focusing on this age range allow us to identify factors associated with school belonging in a critical transition period, where students' sense of school belonging begin to show a general decline (Anderman, 2003). In our models we consider 88 factors at the individual, classroom, peer, teacher, school, family and neighborhood level, which might be associated with school belonging (see, e.g. Allen et al., 2018; Anderman, 2003; Maurizi, Ceballo, Epstein-Ngo, & Cortina, 2013).

Basic principles in statistical learning

Broadly speaking, statistical learning (also called 'machine learning' or 'data mining') refers to a set of methods intended to extract information and knowledge from data (Molina & Garip, 2019). Statistical learning approaches are generally classified in two categories: supervised and unsupervised. The goal of the former is to build a model that accurately predicts an outcome of interest, and the goal of the latter is to uncover patterns among a set of inputs (i.e., no output is available). In this paper, we will focus exclusively on supervised statistical learning (SML) approaches.

The goal of SML is to produce accurate predictions of an outcome Y given a set of inputs X using some function $f(X)$. SML relax some assumptions about the functional form of f, and SML approaches are thus considered non-parametric (Hastie et al., 2009; Kleinberg et al., 2015). For example, rather than assuming that the inputs in f combine in a linear and additive fashion (as in ordinary least squares), SML methods can search for a particular f that maximizes predictive accuracy without assuming linearity or additivity. The fact that $f(X)$ is not constrained by a functional form specified beforehand allows researchers to uncover complex patterns in the data (e.g., related to interactions or non-linearities) that might have been difficult to detect using more rigid model specifications.

Yet the flexibility of SML methods comes with a greater risk of overfitting the data –that is, of following the error, or noise, of the data too closely (James, Witten, Hastie, & Tibshirani, 2013). If the model captures both the signal and the noise of the data at hand, then it will have an excellent fit to the training data (i.e., the data used to estimate $f(X)$). However, we are typically not interested in predicting the scores in the training data (as we know the values of Y for those observations), but rather in predicting the scores of observations we have not yet seen. The goal of SML methods is to achieve an adequate balance between flexibly capturing the patterns in the training data while having a good out-of-sample performance. Put differently, the goal is to estimate a function that minimizes both the in-sample and the out-of-sample prediction error. As suggested above, there is often a tradeoff between these two quantities: as model flexibility increases, the in-sample prediction error (or bias) will decrease; however, if the model is too flexible then it will be too sensitive to the data at hand (i.e., it will vary across training sets), and will not generalize well to other datasets. In the statistical learning literature this is referred to as the bias-variance tradeoff (Hastie et al., 2009).

A key tool employed in SML to address the bias-variance tradeoff is regularization (Berk, 2008; Hastie et al., 2009). The idea behind regularization is to reduce the flexibility of the function (which generates overfitting) by penalizing complexity. Without regularization, the function might have a good in-sample fit but might not generalize well to other samples. By including a regularizer, we can reduce the flexibility of the function to ensure a good out-of-sample performance.

A central part of implementing SML methods consists in choosing the adequate level of regularization. This step is often referred to as 'tunning the algorithm', and is often determined empirically using cross-validation (Berk, 2008; Varian, 2014). The idea behind cross-validation is to estimate the out-of-sample accuracy of the function by fitting the function in one portion of the data (the training set) and evaluating its performance in another portion (the validation set). A common type of cross-validation is k-fold cross validation, which consists in randomly partitioning the sample in k folds, and successively using one of the folds as the validation set and the remaining folds as the test set (Hastie et al., 2009). The best tunning parameter is the one associated with the lowest cross-validated error.

Regularization and empirical tunning are two key elements that distinguish SML from traditional statistical techniques (Mullainathan & Spiess, 2017b). As Mullainathan and Spiess (2017b) explain, any SML algorithm can be defined by a function class F (e.g., trees or smoothing splines), and a regularizer $R(f)$ that determines the complexity of the function. The nature of the regularizer depends on the function class: for example, in the case of trees the complexity might depend on the number of nodes or minimal leaf size, whereas in the case of splines it might depend on the number of knots.

In the present study, we implement two SML methods related to two different classes: the least absolute shrinkage and selection operator (Lasso), which is associated with the class of linear functions; and multivariate adaptive regression splines (MARS), which is associated with the class of splines. These methods differ along several dimensions, notably in terms of their flexibility and interpretability (James et al., 2013). In statistical learning there is often a tradeoff between accuracy and interpretability: more restrictive models tend to be less accurate and more interpretable, while more flexible models tend to be more accurate and less interpretable (see Rudin, 2019). For example, Lasso regression is a restrictive (linear) model which can be easy to interpret, while random forests are highly flexible models that can be difficult to interpret.

The aim of this study is to use SML methods in order to (1) assess the predictability of students' sense of school belonging in fifth grade, and (2) identify the factors that are most predictive. By using two different SML methods, we aim to check whether variable importance holds across model specifications, and whether more flexible models that include complex nonlinearities and interactions have a better performance than less flexible models. We also wish to illustrate how researchers can use SML methods to move beyond the linear-additive models that are commonly used in social and behavioral research. Below, we provide a brief introduction to the two SML methods implemented.

Conducting variable selection with Lasso

A basic premise in social and behavioral research is that human development and behavior is shaped by a multitude of interrelated processes that take place in different contexts (e.g. Bronfenbrenner, 1979; Thelen & Smith, 2007). An implication of this principle is that there can be potentially many variables one can investigate. For example, prior research suggests that a variety of factors can influence school belonging, from individual-level characteristics to peer, family or school-level influences (see, e.g. Allen et al., 2018; Korpershoek et al., 2019). An important task consists in identifying the factors that are most predictive of the outcome of interest –in this case students' sense of school belonging. Rather than selecting a subset of factors based on prior knowledge, as is commonly done in social and behavioral research, we take a data-driven approach, aiming at empirically determining which among a large set of predictors are more influential, and which factors have a negligible association with the outcome.

The Lasso is a well-known statistical learning tool for variable selection. Lasso belongs to the class of linear functions, and is considered a more restrictive approach than OLS, as it sets some of the estimated coefficients to zero (Hastie et al., 2009). By being less flexible, Lasso is also more interpretable than OLS, as it only selects the variables that are most predictive of the outcome. In addition, Lasso can generate more accurate predictions and prevent overfitting due to the inclusion of a regularizer.

The Lasso penalizes complexity by constraining the magnitude of the regression coefficients. This type of regularization is referred to as shrinkage, as the regression coefficients are shrunk toward zero (Berk, 2008). The goal of shrinking the coefficients is to reduces the variance in the predictions. Shrinkage methods are implemented by including a shrinkage penalty when estimating the regression coefficients. More specifically, given a sample of n cases and p predictors, where y_i is the outcome for the i^{th}case and x_{ij} the j^{th} predictor for the i^{th}case, the coefficients in Lasso minimize

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| \\ = RSS+\lambda\sum_{j=1}^{p}|\beta_j|, \quad (1)$$

where $\lambda \geq 0$ is a tunning parameter. Note that the first term is the residual sum of squares (RSS) used to estimate the coefficients in OLS. The second term, $\lambda\sum_{j=1}^{p}|\beta_j|$, is the penalty that shrinks the coefficients toward zero (Hastie et al., 2009). If $\lambda = 0$, then Lasso will produce the OLS estimates. As the value of λ increases, the shrinkage penalty grows, and the coefficients will tend to zero. Variable selection occurs when the coefficient of a set of variables is set to zero. Typically, the value of the tunning parameter λ is empirically determined using cross-validation. The idea behind this approach is to choose different values of λ, compute the cross-validation error for each λ, and then select the value of the tunning parameter with the smallest error (Hastie et al., 2009).

It is worth noting that there are other methods for conducting variable selection. Popular techniques include stepwise methods such as forward and backward stepwise selection, which are fit using least squares. Unlike Lasso regression, these techniques do not use a tuning parameter that determines the complexity of the model. In addition, they are not guaranteed to yield the best model containing a subset of predictors and, contrary to Lasso, require that the number of observations is larger than the number of predictors (James et al., 2013). Even if different techniques can be useful in different scenarios, simulation studies indicate that Lasso outperforms other variable selection methods (Hastie, Tibshirani, & Tibshirani, 2020).

Similarly, it is important to note that apart from Lasso there are other shrinkage methods, e.g., ridge regression (Hastie et al., 2009). Similar to Lasso, ridge regression includes a term in the minimization function, $\lambda\sum_j \beta_j^2$, which shrinks the estimates toward zero, and the tuning parameter can also be determined using cross-validation. By shrinking the coefficients (and, as a consequence, reducing the variance), ridge regression can be used to increase the predictive accuracy of the model. However, ridge regression does not set the coefficients exactly to zero, and as a consequence is not used for variable selection. As explained above, a key goal in this study is to identify the set of variables that are related to our outcome of interest and eliminate all irrelevant predictors (which means setting their coefficient to zero). Thus, we will use Lasso to search for a sparse and interpretable model that includes only a subset of relevant predictors .

Modelling non-linear and non-additive effects with MARS

As explained above, Lasso is a linear regression analysis with a regularizer that can improve prediction accuracy and conduct variable selection. Linear models have several advantages, notably related to inference and interpretability (James et al., 2013). However, some phenomena are inherently nonlinear, so we need to account for nonlinearity in order to adequately represent the phenomena. Researchers often deal with nonlinearities by transforming variables or adding terms to the model (e.g., squared terms). However, in order to do this one needs to know the nature of the nonlinearity in the data, and the results might depend on the particular transformation or terms added to the model –i.e., on the 'researcher degrees of freedom' (Simmons et al., 2011). In order to address this problem, one can use statistical learning tools that estimate nonlinear functions that maximize predictive accuracy.

A SML method that relaxes the linearity assumption while maintaining interpretability is a form of regression called multivariate adaptive regression splines (MARS). MARS accounts for nonlinearities by fitting piecewise linear regressions; that is, by breaking the predictors into different bins, and fitting a linear regression within each of these bins (Hastie et al., 2009; Kuhn & Johnson, 2013). These linear relationships are often called 'hinge' functions (Kuhn & Johnson, 2013). The appropriate number of cut-points (or knots) that determine the number of hinge functions is considered a tuning parameter and can be determined empirically using cross-validation. In this approach, each point for each predictor is evaluated, and the cut-point that achieves the smallest cross-validated error is retained (Kuhn & Johnson, 2013). The algorithm also considers the prediction error of a model without any cut-points. If a predictor (with or without knots) does not improve the prediction accuracy of the model, then it is not retained. Consequently, MARS automatically conducts variable selection.

Apart from modelling nonlinearities, MARS can be used to identify non-additive effects by considering the product of two or more hinge functions. For example, in a second-degree MARS model, second-order interactions between the hinge functions identified are included in the search procedure. The degree of the terms added to the model is considered a tunning parameter, and as a consequence can be determined using cross-validation. In sum, the MARS model has two tunning parameters (the degree of the terms included and the number of terms retained), both of which can be determined using cross-validation (Kuhn & Johnson, 2013).

The MARS model has several advantages. First, as noted above, the model automatically conducts variable selection. Second, the results are highly interpretable (i.e., it is clear how the selected predictors affect the outcome), as each hinge function models in a linear fashion a particular region of the predictor space (Kuhn & Johnson, 2013). Third, MARS models can handle both continuous and categorical data. This is an important advantage in social and behavioral research, where datasets are typically mixed and many of the measures used are ordinal in nature (e.g., they are based on Likert-style response items). Given that Likert-scale items provide ordinal (rather than metric) information, researchers have argued that common descriptive statistics such as means and standard deviations, as well as models that assume a metric scale such as the t-test, analysis of variances and ordinary least-squares regression, should not be used to analyze rating-scale data (Jamieson, 2004; Liddell & Kruschke, 2018). Apart from the ordinal nature of the scale, Likert-style items often violate distributional assumption (e.g., normality) of parametric models, which can systematically lead to errors (Liddell & Kruschke, 2018). The MARS model does not rely on these parametric assumptions, and any nonlinearities are automatically identified by the algorithm. Finally, MARS can be used to estimate the relative importance of the predictors considered by examining the reduction in the prediction error that occurs when adding a particular predictor to the model (for details, see Milborrow, Hastie, & Tibshirani, 2014).

Data and Methods

Analytic Sample

The data comes from the Early Childhood Longitudinal Study (ECLSK: 2010) conducted by the National Center for Education Statistics (see Tourangeau et al., 2018, for more information regarding this study). The study tracks a nationally representative sample of 18,170 U.S. children who entered kindergarten in the 2010–2011 school year through fifth grade. The analytic sample was defined as 11,434 individuals that had a non-missing value in the outcome variable (school belonging in 5th grade).

Missing data

The amount of missing data in the analytic sample was not substantial, as most variables had less than 10% of missing cases, and no variable had more than 30% of missing cases. In order to include the entire sample in the analysis, single imputation methods were conducted for replacing a single value for each missing data point. Longitudinal imputation using the most recent non-missing value was conducted for relatively stable variables (e.g., school district poverty and school safety), and for the remaining variables stochastic regression imputation was conducted. In order to improve the imputation models, several auxiliary variables were added, including the previous value of the imputed variables (if available) and demographic characteristics (Nguyen, Carlin, & Lee, 2017).

Variable selection

Prior research suggests that students' sense of school belonging might be determined by a range of factors at the individual, classroom, peer, teacher, school, family and neighborhood level. For example, prior research suggests that school belonging is associated with a range of student-level characteristics such as students' prosocial behaviors (Demanet & Van Houtte, 2012), psychological well-being (Jose, Ryan, & Pryor, 2012), and involvement in extracurricular activities (Fredricks & Eccles, 2006); teacher-level characteristics such as pedagogical practices (Anderman, 2003) and teacher support (Chiu, Chow, McBride, & Mol, 2016); school-level characteristics such as school size (Anderman, 2003) and school safety (DeRosier & Newcity, 2005); peer-level characteristics such as harassment (Waters, Cross, & Shaw, 2010); and family-level characteristics such as immigrant status and family communication (Chiu et al., 2016).

In the analysis, we considered 89 variables which can be broadly classified in the following categories: children's demographic characteristics (5), cognitive skills (12), socioemotional behaviors (27) and habits (4); family socioeconomic status, composition and dynamics (15); instructional variables related to the curriculum, pedagogy and time spent on different activities (15); and school (10) and neighborhood (1) characteristics (see Table 2 in the Appendix). These variables could also be classified according the that eight general 'themes' that have been found to be associated with school belonging (Allen et al., 2018): academic motivation, parent support, teacher support, emotional stability, peer support, gender, personal characteristics and environmental features. We tried to include all (or at least the majority) of variables in the dataset that could be potentially related to the outcome of interest, while paying attention to the validity and reliability of the constructs . Measurement quality has been generally overlooked in machine learning, and yet it can affect the effectiveness of the algorithms (Jacobucci & Grimm, 2020).

Out of the 89 variables considered, 52 variables were based on Likert-type questions included in the child, teacher, parent or school administrator questionnaire, as well as in the parent interview. These variables were constructed by averaging the items related to the construct. The majority of variables included were measured in fifth grade. If the variable was not measured in fifth grade, then we used its closest value (e.g., if a variable was not measured in fifth grade, then we included the measure obtained in fourth grade; if it was not measured in fourth grade, then we included the measure in third grade, and so on).

The ECLS included 5 Likert-scale items measuring the outcome of interest –students' sense of school belonging– in 5th grade. Students were asked to self-report on a 4-point scale the extent to which they fit in and felt connected with different aspects of the school (Tourangeau et al., 2018). Response options included 1 (*Never*), 2 (*Sometimes*), 3 (*Often*), and 4 (*Always*). The questions asked were: 'This school year, how often did you... (1) Feel like you fit in at your school? (2) Feel close to classmates at your school? (3) Feel close to teachers in your school? (4) Enjoy being at your school? (5) Feel safe at your school?' Individual scores were computed by averaging these 5 items.

Out of the 89 variables considered, 8 variables were categorical. Some of these variables were binary (e.g., disability status and sex) and others had multiple categories (e.g., school type and school location).

Finally, 29 variables were based on a continuous scale. Some of these variables were based on direct assessments (e.g., the variables related to the students' academic achievement and executive functions), while others were based on responses to interviews or questionnaires (e.g., regarding the amount of time the child spends on particular academic and non-academic activities). Table 2 includes all the constructs included in the analysis along with the variable name, dimension, original scale, measurement procedure and measurement occasion (for more information about the variables included, see Tourangeau et al., 2018, and the references therein).

Standardization

Shrinkage methods such as the Lasso regularize the coefficient estimates, and as a consequence the results can vary depending on the scale of the predictors. Consequently, it is recommended to standardize all predictors by dividing each variable by its standard deviation (James et al., 2013). We standardized all predictors following this procedure, and dummy coded all categorical variables. We used this dataset for estimating the Lasso, and the non-standardized dataset for estimating the MARS model.

Training and test datasets

We randomly assigned 9,134 individuals (80%) to a training dataset, and 2,300 individuals (20%) to a test dataset. The training dataset was used for fitting and tuning the models using cross-validation. The test dataset was used to assess the models' final predictive performance and compare the performance across different models. In order to measure the quality of fit we used the models' mean squared error (MSE).

Model selection

All tunning parameters were selected using 10-fold cross-validation applied to the training dataset. We trained the Lasso model using the glmnet package (Friedman, Hastie, & Tibshirani, 2009), and the MARS model using the caret (Kuhn, 2008) and earth (Milborrow et al., 2014) packages. In order to select the most conservative and sparse models, we used the 'one-standard-error' rule, which picks the most parsimonious model within one standard error of the minimum value of the tunning parameter (see Hastie et al., 2009). That is, we select the simplest model that has a similar predictive accuracy to the best model. The reason for doing this is that we valued interpretability in addition to predictive accuracy.

Results

Baseline OLS regression

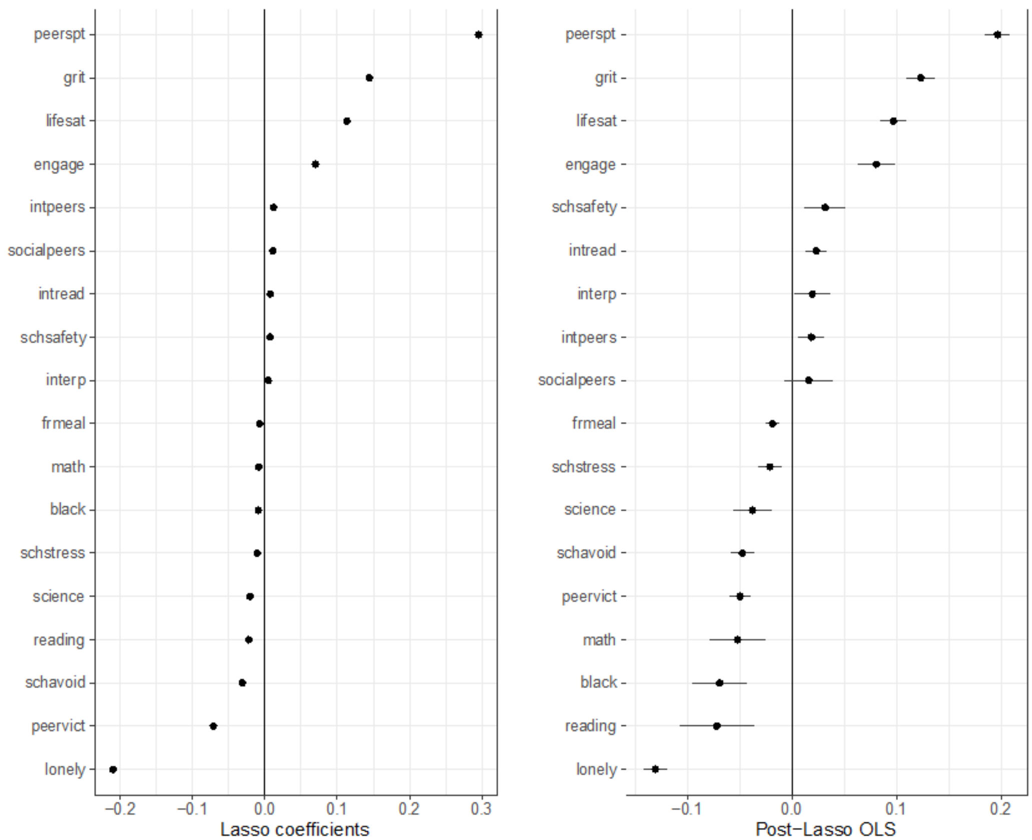
We began by fitting a linear model using OLS in the training data with all 88 predictors, and estimated the models' predictive accuracy using the test data. The MSE in the training data was 0.154 and the R-squared was 0.520. On the other hand, the MSE in the test data was 0.168 and the R-squared was 0.508. The decrease in the predictive accuracy in the test set compared to the training set suggests that some overfitting has occurred.

By estimating 88 coefficients, the linear model using OLS is difficult to interpret and unnecessarily complex. In order to remove the irrelevant variables, we can perform variable selection using techniques such as Lasso regression.

Lasso regression

In order to select the tunning parameter λ, we examined the cross-validated mean-squared error across the entire range of possible solutions and selected the 'best' model, defined as the most parsimonious and accurate model using the 'one-standard-error' rule (Hastie et al., 2009). The best model had a value of $\lambda = 0.023$. Due to variable standardization, interpreting the Lasso results presents some difficulties. In particular, the MSE is scale-dependent, so we cannot use it to compare the predictive accuracy with models that do not require standardization. In addition, it is often preferable to interpret the estimated coefficients in their original metric. Consequently, we conducted a post-Lasso OLS, where the variables selected by Lasso were

Figure 1 ■ Lasso and Post-Lasso OLS coefficients. The Lasso coefficients were estimated using the training dataset and standardized covariates. The Post-Lasso OLS were estimated using the entire dataset and non-standardized covariates.



used to fit an OLS regression (for a discussion of this approach, see Belloni & Chernozhukov, 2013).

As Table 1 indicates, the predictive accuracy on the test set of the baseline OLS model (MSE = 0.168) is similar to the predictive accuracy of the post-Lasso OLS model (MSE = 0.170) . This suggests that the Lasso procedure adequately selected the variables that are most predictive of students' sense of school belonging. In particular, the Lasso indicates that 18 predictors have a non-zero coefficient. The predictors retained by the model along with their associated coefficients are depicted on the left panel in Figure 1. One can see that the largest coefficient (0.296) is associated with $peerspt$, which represents students' feelings of peer social support . The second largest coefficient (-0.210) is associated with *lonely*, which represents students' feelings of loneliness at school . One can also see that students' grit and life satisfaction have a standardized coefficient above 0.1, and students' behavioral engagement and peer victimization have an associated coefficient above 0.05.

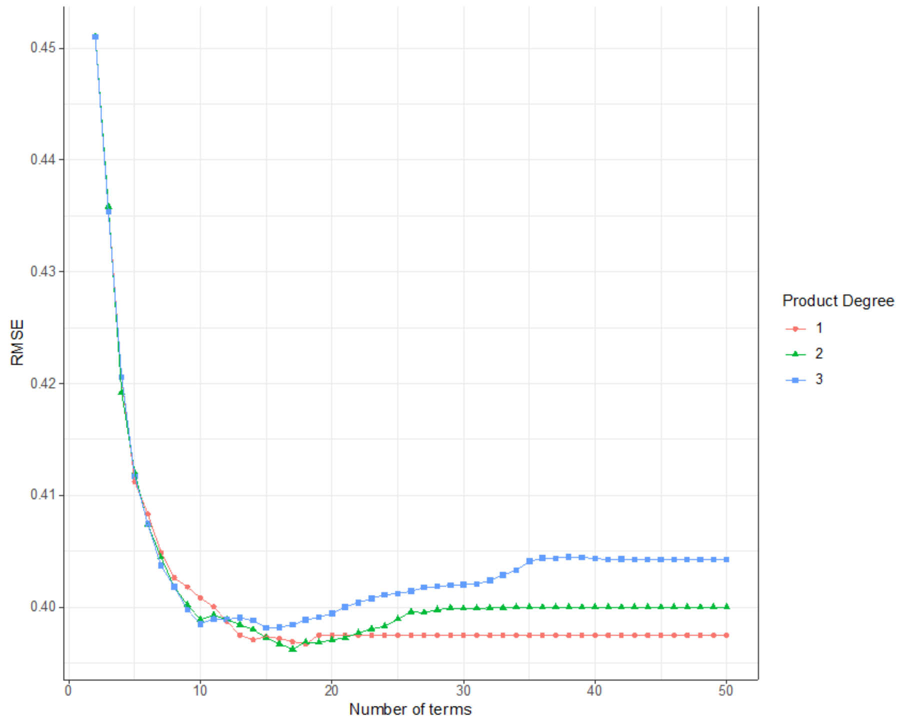
The right panel in Figure 1 depicts the post-Lasso OLS coefficients along with their confidence intervals. One can see that the Lasso coefficients tend to be shrunken with respect to the OLS coefficients. One can also see that the rank order of the strongest predictors remains the same –in particular, the strongest predictor remains $peerspt$ (0.19) fallowed by *lonely* (-0.13). It is worth noting that the goal of SML procedures is to produce accurate predictions of an outcome variable rather than generating unbiased parameter estimates (Mullainathan & Spiess, 2017a). It is important, then, to not draw any causal inferences from these estimated coefficients and their confidence intervals.

MARS regression

There are two tunning parameters in the MARS model: the number of terms retained and the degree of interactions. As explained above, a term can be composed of a single variable or a hinge function. In order to tune the algorithm, we conducted a grid search to identify the optimal combination of these tunning parameters. As with the other techniques implemented in this study, we used the

2021 ■ Vol. 17 ■ no. 3

Figure 2 ■ Cross-validated RMSE for MARS models with different combinations of the two tunning parameters (the degree of interactions and the number of terms). The minimum RMSE is associated with a model with 16 terms and second-degree interactions. We selected a model with 8 terms and no interactions, following the one-standard-error rule.



models' cross-validated prediction error to select the best model, defined as the simplest model within one standard deviation of the model that minimizes the cross-validated prediction error. The grid search was specified as 147 possible combinations of the degree of interactions (first, second, and third degree) and the number of terms retained (2 to 50).

The results of the grid search are depicted in Figure 2. The figure shows the cross-validated root-mean-square error (RMSE) of all possible combinations. One can see that the minimum RMSE is associated with a model of 17 terms with second degree interactions (which we report below). We selected the simplest model within one standard deviation of this model. Our final model retained 12 terms composed of 8 variables and no interactions. The model can be written as

$$\begin{aligned}\widehat{schbelong} = 3.14 &+ 0.27\max(0, 1.11 - read) - 0.15\max(0, 4.0 - grit) + 0.11\max(0, grit - 4.0)\\ &+0.27\max(0, 1.33 - lonely) - 0.13\max(0, lonely - 1.33) - 0.20\max(0, 4.67 - peerspt)\\ &+0.27\max(0, peerspt - 4.67) + 0.15\max(0, engage - 4.20) + 0.22\max(0, lifesat - 4.33)\\ &+0.09\max(0, 2.25 - peervict) + 0.06\max(0, 2.83 - scavoid).\end{aligned} \tag{2}$$

This equation shows how the MARS algorithm accounts for non-linear relationships by including hinge functions while preserving an additive structure. For example, the term $0.22\max(0,\ lifesat - 4.33)$ implies that a one-unit difference in $lifesat$ is associated with a 0.22 difference in $\widehat{schbelong}$ when $lifesat\ > 4.33$.

Table 1 ■ Fit statistics across models

	Training set		Test set	
	MSE	R^2	MSE	R^2
Baseline OLS	0.154	0.520	0.168	0.508
Post-Lasso OLS	0.156	0.512	0.170	0.504
MARS	0.158	0.506	0.172	0.496

Note. The training set was used to fit all models using 10-fold cross validation, and the test set was used to estimate the generalization error. The training set consists of 9,134 (≈80%) randomly assigned individuals, and the test set includes the remaining 2,300 (≈20%).

Given the number of terms involved, it is difficult to perceive how each hinge function contributes to the model. One way of visualizing the independent effect of each predictor is by constructing the prediction profile or partial dependence plot of each variable holding the other variables at their mean level (Kuhn & Johnson, 2013). Figure 3 illustrates the prediction profile of the 8 variables retained in the model. The figure suggests that some variables might have non-linear relationships with the outcome, e.g., $engage$, $schavoid$, $peervict$, $lifesat$ and $reading$. In particular, changes in $peervict$, $schavoid$ and $reading$ appear to be more predictive in the lower end of the scale, whereas changes in $lifesat$ and $engage$ appear to be more predictive in the upper end of the scale. On the other hand, other variables such as $grit$, $lonely$ and $peerspt$ (the strongest predictors of $schbelong$) appear to have a fairly linear relationship with the outcome.

Figure 4 depicts the partial dependence plots of the MARS model with the minimum RMSE. The model includes 17 terms composed of 12 different variables. The model includes 9 interactions between hinge functions. Interestingly, 5 of these interaction terms include behavioral engagement (two interactions with $grit$ and $schavoid$, and one interaction with $hmwkread$). Figure 4 shows how school belonging decreases rapidly when low levels of school engagement combine with these other factors. Even if these interactions can be informative and can be investigated in future research, they are presented here only for illustrative purposes. The best fitting model is not only more complex, but also more likely to overfit, and as a consequence should be interpreted with caution. One would be more confident about the presence of these interactive effects if they were also found in the parsimonious model (especially if they involved the strongest predictors). From now on, we will focus then on the parsimonious model.

Another way of interpreting the MARS model is by quantifying the importance of each of the predictors selected by the algorithm. This is generally done by examining the reduction in the prediction error that occurs when including the predictor in the model (Kuhn & Johnson, 2013). For example, Figure 2 indicates that by adding two predictors the RMSE is reduced from around 0.45 to around 0.42. Figure 5 presents the importance scores associated to each variable, scaled between 0 and 100. The prediction error considered is the residual sum-of-squares. Consistent with prior results, $peerspt$ appears to be the most important variable for predicting students' sense of school belonging, followed by $lonely$ and $lifesat$. On the other hand, $schavoid$, and $reading$ seem to have a negligible predictive power in this particular model.

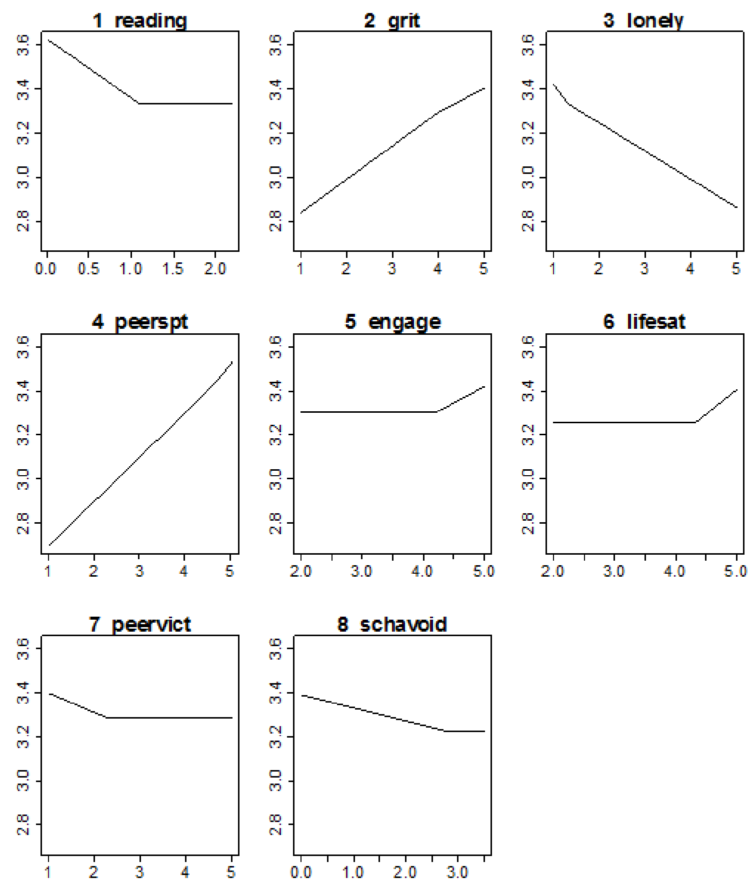
Finally, Table 1 indicates that the MARS model selected does not improve the predictive performance compared to the two linear models estimated before. Yet it is worth noting that the 'best' MARS model was selected not only based on its predictive accuracy but also on its simplicity. Overall, however, the results indicate that a linear model provides a good approximation of the relationships between the predictors and the outcome of interest.

Even if in this particular case fitting a nonlinear model might not be necessary, it is important to note that it is difficult to establish beforehand the presence or absence of nonlinearities. Thus, researchers recommend allowing for nonlinearities (especially for important predictors), and implement linear models only if nonlinear effects are not detected (Harrell, 2015). Finally, the fact that a linear model fits the data well might be due to the nature of the phenomena represented, but also to limitations in the data (as most variables are based on 4 to 6-point Likert scale items, which might limit the ability to identify nonlinearities).

Discussion

The sense of school belonging refers to students' feelings of being accepted and connected to their particular school (Anderman, 2003). School belonging has been considered an important determinant of a range of academic and socioemotional outcomes. For example, prior researcher suggests that school belonging is negatively related to students' depression and anxiety (Shochet, Dadds, Ham, & Montague, 2006), school dropout (Archambault, Janosz, Fallu, & Pagani, 2009) and general feelings of alienation (Hascher & Hadjar, 2018). Yet despite an extensive liter-

Figure 3 ■ Partial dependence plots of the eight variables in the MARS model. Each plot depicts the estimated relationship between the predictor and the outcome while holding the other predictors at their mean value.

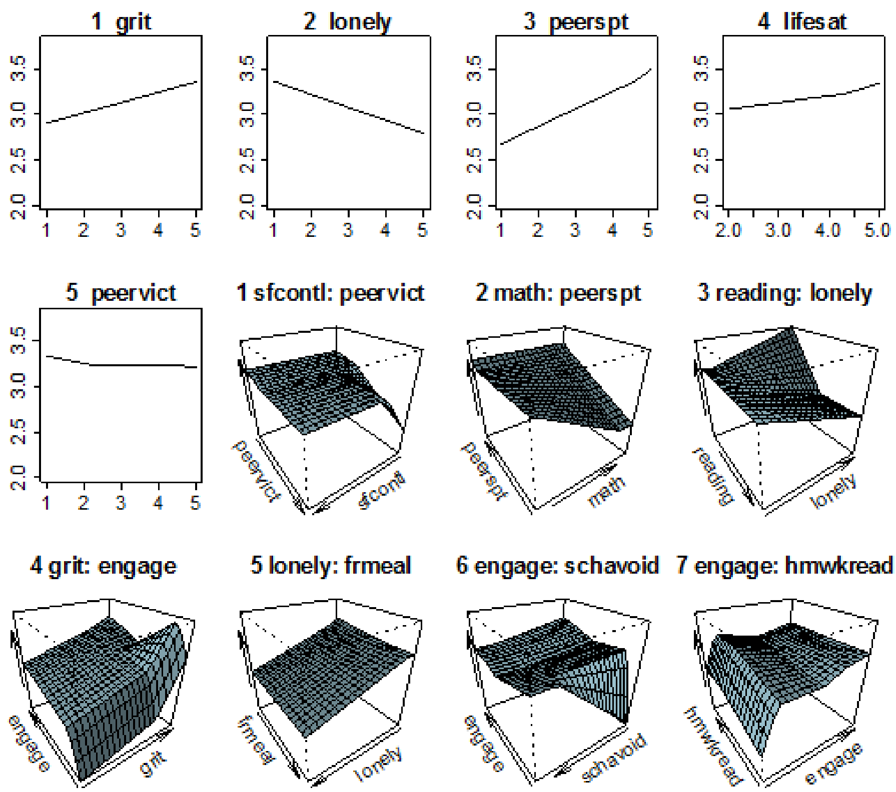


ature on the topic, it is not clear what factors are more strongly related to the students' sense of school belonging. Prior research has identified a wide range of individual and contextual level factors that might be related to school belonging, from students' psychological well-being and involvement in extracurricular activities, to various teacher, family and school level characteristics (e.g. Anderman, 2003; Chiu & Xu, 2020; DeRosier & Newcity, 2005; Fredricks & Eccles, 2006; Jose et al., 2012).

In the present study, we investigated the extent to which students' sense of school belonging can be predicted by a wide range of individual and contextual-level factors using two statistical learning techniques (Lasso and MARS). The results suggest that the predictive accuracy of the SML algorithms is comparable to the predictive accuracy generated by a linear model estimated by OLS using all 88 predictors. The estimated test root-mean-square error of the three models was around 0.41, which means that our predictions would be off by around 0.41 units in the original scale. It is also worth noting that the test R-square was around 0.50, which indicates that 50% of the variance in school belonging can be explained by the models. How accurate or useful are these predictions would depend on how the predicted values are used.

One of the main advantages of the SML methods implemented in this study is that they allow us to generate accurate predictions while performing variable selection. Thus, we are able to approximate the predictive accuracy of the baseline OLS model that includes 88 covariates with models that include only 18 (Lasso) or 8 (MARS) predictors. Creating models that are interpretable is a fundamental value in scientific research as well as in policy design and implementation (e.g. Rudin, 2019). This is the reason why we not only considered prediction accuracy but also sparsity

Figure 4 ■ Partial dependence plots of the MARS model with the lowest RMSE.



for model selection. The selected models and results are, then, interpretable and useful for theory and hypothesis generation.

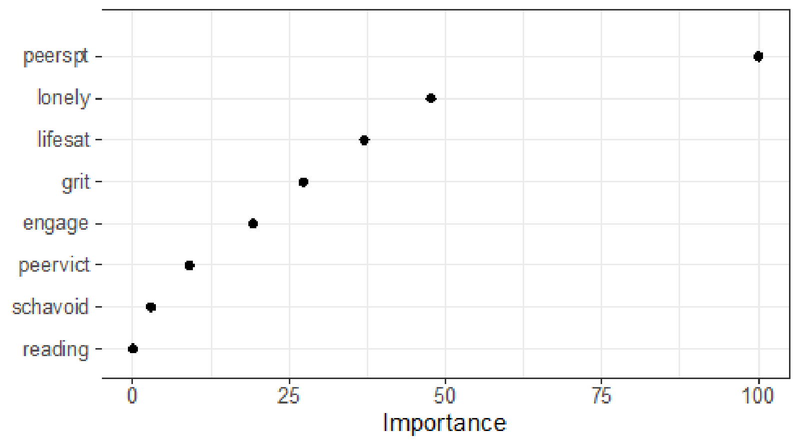
Apart from selecting a subset of predictors, we used the SML techniques to estimate which variables are more strongly related to students' sense of school belonging. The results across all model specifications indicate that students' feelings of peer social support are the most predictive factor, followed by students' feelings of loneliness at school. This suggests that peer relationships are a central factor related to students' sense of school belonging. In addition to these social components, students' grit and life satisfaction are also predictive of students' sense of belonging in school. These findings appear to contradict previous research suggesting that teacher-level predictors (in particular teacher support) are the strongest predictors of school belonging (Allen et al., 2018). However, it is important to note that our dataset does not include all relevant predictors, including some of the teacher-level predictors that have been found to be important determinants of students' feelings of school belonging.

Finally, we used SML techniques to consider whether complex functions including non-linear and interactive effects produced more accurate predictions or substantially different results. Many phenomena in social and behavioral sciences cannot be properly represented using linear and additive models (e.g. Braumoeller, 2003; Ragin, 2009). It is important, then, to consider whether more complex relationships provide a better fit to the data. Model-based findings often depend on researchers' decisions to impose a particular functional form or include a particular set of covariates. These 'researcher degrees of freedom' can generate false positives or non-replicable results (Gelman & Loken, 2013; Simmons et al., 2011). An advantage of SML models is that one can automatically search for complex nonlinear and interaction effects while having an objective evaluation criterion.

References

Allen, K., Kern, M. L., Vella-Brodrick, D., Hattie, J., & Waters, L. (2018). What schools need to know about fostering school belonging: A meta-analysis. *Educational Psychology Review*, *30*, 1–34.

Figure 5 ■ Predictor importance plot based on the MARS model. This measure reflects the change in residual sums of squares as each predictor is added to the model. Given that this is a relative measure, it is normalized so the largest decrease is 100.



Anderman, L. H. (2003). Academic and social perceptions as predictors of change in middle school students' sense of school belonging. *The Journal of Experimental Education*, *72*(1), 5–22.

Anderman, L. H., & Freeman, T. M. (2004). Students' sense of belonging in school. *Advances in motivation and achievement*, *13*, 27–63.

Archambault, I., Janosz, M., Fallu, J. S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of adolescence*, *32*, 651–670.

Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, *19*(2), 521–547.

Berk, R. A. (2008). *Statistical learning from a regression perspective (vol. 14)*. New York: Springer.

Braumoeller, B. F. (2003). Causal complexity and the study of politics. *Political Analysis*, *11*(3), 209–233.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge. MA: Harvard University Press.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press.

Chiu, A., & Xu, Y. (2020). Bayesian rule set: A quantitative alternative to qualitative comparative analysis. Retrieved August 16, 2021, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3639664

Chiu, M. M., Chow, B. W. Y., McBride, C., & Mol, S. T. (2016). Students' sense of belonging at school in 41 countries: Cross-cultural variability. *Journal of Cross-Cultural Psychology*, *47*, 175–196.

Demanet, J., & Van Houtte, M. (2012). School belonging and school misconduct: The differing role of teacher and peer attachment. *Journal of youth and adolescence*, *41*, 499–514.

DeRosier, M. E., & Newcity, J. (2005). Students' perceptions of the school climate: Implications for school safety. *Journal of School Violence*, *4*(3), 3–19.

Fredricks, J. A., & Eccles, J. S. (2006). Is extracurricular participation associated with beneficial outcomes? concurrent and longitudinal relations. *Developmental Psychology*, *42*(4), 698–713.

Fried, E. I., & Robinaugh, D. J. (2020). Systems all the way down: Embracing complexity in mental health research. *BMC Medicine*, *18*(1), 205–205. doi:10.1186/s12916-020-01668-w

Friedman, J., Hastie, T., & Tibshirani, R. (2009). Glmnet: Lasso and elastic-net regularized generalized linear models. *R Package Version*, *1*(4), 1–24.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Retrieved August 16, 2021, from http://www.stat.columbia.edu/gelman/research/unpublished/p_hacking.pdf

Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal re-*

gression, and survival analysis. doi:10.1007/978-3-319-19425-7

Hascher, T., & Hadjar, A. (2018). School alienation–theoretical approaches and educational research. *Educational Research*, *60*, 171–188.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Science & Business Media: Springer.

Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons. *Statistical Science*, *35*(4), 579–592. doi:10.1214/19-STS733

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, *355*(6324), 486–488.

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. doi:10.1177/1745691620902467

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning (vol. 112)*. New York: Springer.

Jamieson, S. (2004). Likert scales: How to (ab) use them? *Medical Education*, *38*(12), 1217–1218.

Jose, P. E., Ryan, N., & Pryor, J. (2012). Does social connectedness promote a greater sense of well-being in adolescence over time? *Journal of Research on Adolescence*, *22*(2), 235–251.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, *105*(5), 491–495.

Korpershoek, H., Canrinus, E. T., Fokkens-Bruinsma, M., & de Boer, H. (2019). *The relationships between school belonging and students' motivational, social-emotional, behavioural, and academic outcomes in secondary education: A meta-analytic review*.

Kuhn, M. (2008). Building predictive models in r using the caret package. *J Stat Softw*, *28*(5), 1–26.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling (vol. 26)*. New York: Springer.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

Maurizi, L. K., Ceballo, R., Epstein-Ngo, Q., & Cortina, K. S. (2013). Does neighborhood belonging matter? examining school and neighborhood belonging as protective factors for latino adolescents. *American Journal of Orthopsychiatry*, *83*(2–3), 323–334. doi:10.1111/ajop.12017

Milborrow, S., Hastie, T., & Tibshirani, R. (2014). Earth: Multivariate adaptive regression spline models [r package] (Version 3). Retrieved from http://cran.rproject.org/web/packages/earth/index.html

Molina, M., & Garip, F. (2019). Machine learning for sociology. *Annual Review of Sociology*, *45*, 27–45.

Mullainathan, S., & Spiess, J. (2017a). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Mullainathan, S., & Spiess, J. (2017b). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Nguyen, C. D., Carlin, J. B., & Lee, K. J. (2017). Model checking in multiple imputation: An overview and case study. *Emerging Themes in Epidemiology*, *14*(1), 8–99.

Ragin, C. C. (2009). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago: University of Press.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310.

Shochet, I. M., Dadds, M. R., Ham, D., & Montague, R. (2006). School connectedness is an underemphasized parameter in adolescent mental health: Results of a community prediction study. *Journal of Clinical Child & Adolescent Psychology*, *35*, 170–179.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Thelen, E., & Smith, L. B. (2007). *Dynamic systems theories*.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., & Najarian, M. (2018). Early childhood longitudinal study, kindergarten class of 2010-11 (ecls-k: 2011): User's manual for the ecls-k: 2011 kindergarten-fourth grade data file and electronic codebook, public version, Washington.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28.

Waters, S., Cross, D., & Shaw, T. (2010). Does the nature of schools matter? An exploration of selected school ecology factors on adolescent perceptions of school connectedness. *British Journal of Educational Psychology*, *80*, 381–402.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.

Appendix A

Table 2 ■ Description of the eighty-nine variables included in the analysis.

Abbreviation	Construct	Dimension	Original scale	Measurement procedure[1]	Measurement occasion (grade)
age	Age at assessment	Demographic	Continuous	Collected from administrative data and child assessments	5th
bmi	Body mass index	Demographic	Continuous	Calculated by ECLS	5th
disability	Disability status	Demographic	Categorical	Collected from PI	5th
sex	Sex	Demographic	Categorical	Collected from school and PI	5th
race	Race/ethnicity	Demographic	Categorical	Collected from PI and the Field Management System	5th
approach	Approaches to learning	Cognitive	Ordinal	7 items from TQ	5th
attention	Attentional focusing	Cognitive	Continuous	6 items from TQ	5th
control	Inhibitory control and attention	Cognitive	Continuous	Task composed of 20 trials	5th
grit	Perseverance over the long term	Cognitive	Ordinal	6 items from CQ	5th
intmath	Perceived interest/competence in mathematics	Cognitive	Ordinal	5 items from CQ	3rd
intread	Perceived interest/competence in reading	Cognitive	Ordinal	5 items from CQ	3rd
intscience	Perceived interest/competence in science	Cognitive	Ordinal	5 items from CQ	3rd
math	Knowledge and skills in mathing in mathematics	Cognitive	Continuous	Mathematics assessment	5th
reading	Knowledge and skills in reading in reading	Cognitive	Continuous	Reading assessment	5th
science	Knowledge and skills in reading in science	Cognitive	Continuous	Science assessment	5th
workmem	Working memory	Cognitive	Continuous	Task composed of 30 items	5th
cogflex	Cognitive flexibility	Cognitive	Continuous	Task composed of 40 trials	5th
schbelong	School belonging	Socioemotional	Ordinal	5 items from CQ	5th
engage	Behavioral engagement	Socioemotional	Ordinal	5 items from CQ	5th
externalize	Externalizing problem behaviors	Socioemotional	Ordinal	6 items from TQ	5th
incontrol	Inhibitory control	Socioemotional	Ordinal	7 items from TQ	5th
internalize	Internalizing problem behaviors	Socioemotional	Ordinal	4 items from TQ	5th
interp	Interpersonal skills	Socioemotional	Ordinal	5 items from TQ	5th
intpeers	Perceived interest/competence in peer relationships	Socioemotional	Ordinal	6 items from CQ	3rd
lifesat	Life satisfaction	Socioemotional	Ordinal	6 items from CQ	5th
lonely	Loneliness	Socioemotional	Ordinal	3 items from CQ	5th

Abbreviation	Construct	Dimension	Original scale	Measurement procedure[1]	Measurement occasion (grade)
peervict	Peer victimization	Socioemotional	Ordinal	4 items from CQ	5th
prosocial	Prosocial behavior	Socioemotional	Ordinal	3 items from CQ	3rd
socanxiety	Social anxiety/Fear of negative evaluation	Socioemotional	Ordinal	3 items from CQ	5th
schlike	School liking	Socioemotional	Ordinal	7 items from TQ	5th
sfcontl	Self-control	Socioemotional	Ordinal	4 items from TQ	5th
socialpeers	Prosocial with peers	Socioemotional	Ordinal	5 items from TQ	3rd
socskills	Social skills with peers	Socioemotional	Ordinal	4 items from TQ	5th
peerspt	Peer social support	Socioemotional	Ordinal	6 items from CQ	5th
strucplay	Physical activity during structured play time	Socioemotional	Ordinal	1 item from TQ	3rd
unstplay	Physical activity during unstructured play time	Socioemotional	Ordinal	1 item from TQ	3rd
schstress	Worry/stress about school	Socioemotional	Ordinal	5 items from CQ	5th
tconflict	Conflict with teacher	Socioemotional	Ordinal	8 items from TQ	3rd
tclose	Closeness with teacher	Socioemotional	Ordinal	7 items from TQ	3rd
aggressor	Peer victimization (child as aggressor)	Socioemotional	Ordinal	4 items from TQ	5th
victim	Peer victimization (child as victim)	Socioemotional	Ordinal	4 items from TQ	5th
excluded	Excluded by peers	Socioemotional	Ordinal	4 items from TQ	5th
schavoid	School avoidance	Socioemotional	Ordinal	5 items from PQ	5th
vidgame	Time spent playing videogames	Habit	Continuous	1 item from PQ	3rd
tvtime	Time spent watching TV	Habit	Continuous	1 item from PQ	3rd
mediause	Media usage	Habit	Ordinal	3 items from CQ	5th
exercise	Time spent exercising	Habit	Continuous	1 item from PQ	3rd
homelit	Home literacy environment	Family	Continuous	3 items from PQ	2nd
housetotal	Total number of household members	Family	Continuous	Collected from PI	5th
pincome	Household income	Family	Continuous	Collected from PI	5th
parented	Parent education level	Family	Ordinal	Collected from PI	5th
noneng	Primary language in the child's home	Family	Categorical	Collected from PI	2nd
numbooks	Number of books the child has	Family	Continuous	1 item from PQ	3rd
pstrain	Parental strain	Family	Ordinal	4 questions from PI	1st
pcomm	Parent-child communication	Family	Ordinal	6 questions from PI	3rd
pdepress	Parental depression	Family	Ordinal	12 questions from PI	5th
pexpect	Parental academic expectations of the child	Family	Ordinal	1 question from PI	3rd
pwarm	Parental warmth	Family	Ordinal	4 questions from PI	3rd
foodsecty	Household's food security status	Family	Ordinal	10 questions from PI	5th
spank	Frequency in which the parent spanked the child	Family	Ordinal	1 item from PQ	Kindergarten

Abbreviation	Construct	Dimension	Original scale	Measurement procedure[1]	Measurement occasion (grade)
pdivorce	Parental divorce or separation	Family	Categorical	1 item from PQ	5th
monitoring	Parental monitoring	Family	Ordinal	3 items from CQ	5th
famstruc	Types of parents in the household	Family	Categorical	Collected from PI	5th
flldirections	Teacher's attention to the child's direction-following ability	Instruction	Ordinal	1 item from TQ	5th
timeart	Time spent in art, music and physical education	Instruction	Continuous	6 items from TQ	2nd
timemath	Time spent in mathematics	Instruction	Continuous	1 item from TQ	2nd
timeread	Time spent in reading and language arts	Instruction	Continuous	1 item from TQ	2nd
timescisoc	Time spent in science and social studies	Instruction	Continuous	2 items from TQ	2nd
freetime	Time for unstructured activities (playing and lunch)	Instruction	Continuous	2 items from TQ	3rd
hmwkread	Time dedicated to homework	Instruction	Continuous	2 items from TQ	3rd
indepwork	Time spent on individual work	Instruction	Continuous	4 items from TQ	2nd
tchcenter	Time spent on teacher-centered instruction	Instruction	Ordinal	1 item from TQ	2nd
schrecess	Time dedicated to recess	Instruction	Continuous	2 items from TQ	3rd
stdtests	Use of state or local standardized tests	Instruction	Ordinal	1 item from TQ	2nd
skilsmath	Coverage of mathematics skills	Instruction	Continuous	32 items from TQ	2nd
skilsread	Coverage of reading skills	Instruction	Continuous	33 items from TQ	2nd
quizzes	Use of classroom tests or quizzes	Instruction	Ordinal	1 item from TQ	2nd
tchobs	Focus on students' mastery of objectives or standards	Instruction	Ordinal	1 item from TQ	2nd
edstand	Evaluation based on standards	Instruction	Ordinal	1 item from TQ	5th
blwmath	Students below grade level in mathematics	School	Continuous	1 item from TQ	5th
blwread	Students below grade level in reading	School	Continuous	1 item from TQ	5th
schsafety	School safety	School	Ordinal	6 items from SAQ	5th
schnbhd	School neighborhood safety	School	Ordinal	6 items from SAQ	5th
nonwhite	Percent of non-white students in the school	School	Ordinal	Collected from SAQ	5th
frmeal	Percent of students in the school approved for free or reduced-price meals	School	Ordinal	Collected from SAQ	5th
schtype	School type	School	Categorical	Collected from SAQ	5th
location	School locality	School	Categorical	Provided by ECLS	5th

Abbreviation	Construct	Dimension	Original scale	Measurement procedure[1]	Measurement occasion (grade)
distpov	School district poverty	School	Continuous	Collected from SAQ	5th
nbhdsafe	House neighborhood safety	Neighborhood	Ordinal	3 items from PQ	Kindergarten

Note: [1]: CQ = child questionnaire; TQ = teacher questionnaire; PQ = parent questionnaire; PI = parent interview; SAQ = school administrator questionnaire.

Open practices

The *Open Material* badge was earned because supplementary material(s) are available on the journal's web site.

Citation

Quintana, R. (2021). Who belongs in school? Using statistical learning techniques to identify linear, nonlinear and interactive effects. *The Quantitative Methods for Psychology*, *17*(3), 312–328. doi:10.20982/tqmp.17.3.p312

Received: 09/06/2021 ∼ Accepted: 16/08/2021