DOI: 10.1111/cdev.13910

EMPIRICAL ARTICLE

CHILD DEVELOPMENT

Cheating among elementary school children: A machine learning approach

Li Zhao[1] | Yi Zheng[2] | Junbang Zhao[3] | Guoqiang Li[4] | Brian J. Compton[5] | Rui Zhang[6] | Fang Fang[7,8,9,10] | Gail D. Heyman[5] | Kang Lee[11]

[1]Department of Psychology, Hangzhou Normal University, Hangzhou, China
[2]Jing Hengyi School of Education, Hangzhou Normal University, Hangzhou, China
[3]College of Child Development and Education, Zhejiang Normal University, Hangzhou, China
[4]Jing Hengyi School of Education, Hangzhou, China
[5]Department of Psychology, University of California San Diego, San Diego, California, USA
[6]Hangzhou Xiayan Elementary School, Hangzhou, China
[7]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China
[8]IDG/McGovern Institute for Brain Research, Peking University, Beijing, China
[9]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China
[10]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China
[11]Ontario Institute for Studies in Education, University of Toronto, Toronto, Ontario, Canada

Correspondence
Li Zhao, Department of Psychology, Hangzhou Normal University, No. 2318 Yuhangtang Rd, Yuhang District, Hangzhou 311121, China.
Email: zhaoli@hznu.edu.cn

Fang Fang, School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, No. 5 Yiheyuan Rd, Haidian District, Beijing 100871, China.
Email: ffang@pku.edu.cn

Kang Lee, Ontario Institute for Studies in Education, University of Toronto, 45 Walmer Rd, Toronto, ON M5R 2X2, Canada.
Email: kang.lee@utoronto.ca

Funding information
Major Project of Philosophy and Social Science Research of the Ministry of Education of China, Grant/Award Number: 22JZD044; National Natural Science Foundation of China, Grant/Award Number: 32171060; Scientific Research Fund of Zhejiang Provincial Education Department of China, Grant/Award Number: Y202250508

Abstract
Academic cheating is common, but little is known about its early emergence. It was examined among Chinese second to sixth graders (N = 2094; 53% boys, collected between 2018 and 2019) using a machine learning approach. Overall, 25.74% reported having cheated, which was predicted by the best machine learning algorithm (Random Forest) at a mean accuracy of 81.43%. Cheating was most strongly predicted by children's beliefs about the acceptability of cheating and the observed prevalence and frequency of peer cheating at school. These findings provide important insights about the early development of academic cheating, and how to promote academic integrity and limit cheating before it becomes entrenched. The present research demonstrates that machine learning can be effectively used to analyze developmental data.

Abbreviations: AUC, area under curve; GEE, Generalized Estimating Equation; GLM, Generalized Linear Modeling; LR, logistic regression; MLP, Multilayer Perceptron; RF, Random Forest; SES, socioeconomic status; XGBoost, Extreme Gradient Boosting.

Co-first author: Y. Zheng.

INTRODUCTION

Cheating is a common human behavior. In childhood, it often takes the form of academic cheating (Waltzer & Dahl, 2020), which we define as an intentional act performed surreptitiously and illegitimately for the purpose of achieving a desired academic outcome. Academic cheating has serious negative consequences because it devalues learning and undermines perceived fairness (Zhao et al., 2020). The academic cheating cases that receive media attention appear to be just the tip of the iceberg, given the extensive evidence that academic cheating is widespread around the world (Anderman & Midgley, 2004; Cizek, 1999; Cuadrado et al., 2019; Ghanem & Mozahem, 2019; Hrabak et al., 2004). Although prior research on this topic has advanced our understanding of this phenomenon among high school and university students, very little is known about the emergence of academic cheating among younger students. The present study seeks to fill this gap in the literature by using advanced machine learning approach to identify the key factors that contribute to cheating on exams by elementary school children.

Scientific research on academic cheating is extensive, dating back to the 1920s (Hartshorne & May, 1928; Voelker, 1921). Based on the existing evidence, Whitley (1998) proposed an influential model that identifies students as motivated to cheat based on three factors: (1) attitudes about cheating-related social norms and moral obligations, (2) the expected benefits of cheating, and (3) the perceived risk of being caught.

A large number of empirical studies have subsequently confirmed the validity of this model among high school and university students. For example, a recent meta-analysis confirmed that students' attitudes about academic cheating (Cohen's d = 0.77; Lee et al., 2020), and their views about the acceptability of academic misconducts in particular (Ives & Giukin, 2020; Özcan et al., 2019; Cohen's d = 0.98–1.42), are significantly associated with college students' academic dishonesty. This meta-analysis (Lee et al., 2020) also found that students who care a great deal about good grades, and thus stand to benefit more from cheating, are more likely to cheat (Cohen's d = 0.58). In addition, many studies have found that students with poor performance are more inclined to cheat than those with strong performance (Özcan et al., 2019; Cohen's d = 0.28). Regarding the perceived risk of being caught cheating, McCabe and Abdallah (2008) found that college students were more likely to cheat if they reported that they have frequently observed their peers' cheating (Cohen's d = 0.68). One reason may be that observing peers successfully engage in cheating leads students to infer that cheating is not an especially risky thing to do (Murdock & Anderman, 2006). Furthermore, the severity of consequences for cheating affects students' perception of risk (Megehee & Spake, 2008; Cohen's d = 0.51) and students tend to consider sanctions against cheating to be a highly effective way to prevent it (Miller et al., 2011; Siniver, 2013; Cohen's d = 0.36–0.61).

Previous research on academic cheating among high school and university students has revealed other factors associated with academic cheating. One such factor is grade (Whitley, 1998; Cohen's d = 0.56): cheating increases with grade during the middle school to high school years, and then stabilizes or declines slightly as students make the transition to university (e.g., Błachnio, 2019; Desalegn & Berhan, 2014; Hrabak et al., 2004; Turnipseed & Landay, 2018; see Cizek, 1999; Whitley, 1998 for reviews). Gender might matter too: there is evidence that male students cheat more often than female students, but the effect size varies considerably across samples (Newstead et al., 1996; Whitley, 1998; Yu et al., 2016; see Whitley et al., 1999 for a meta-analysis; Cohen's d = 0.15–0.39). Factors such as the family's socioeconomic status (SES; Alt, 2015; Yu et al., 2016; Cohen's d = 0.20–0.22) and parent–child relationships (Bong, 2008; Cohen's d = 0.19) have small effects on students' academic cheating.

Although the first systematic study of academic cheating involved children as young as fourth grade (Hartshorne & May, 1928), only a small number of subsequent studies on cheating among young children has focused on academic cheating (e.g., Zhao et al., 2018, 2020, 2021). Most of this work has instead focused on cheating to win material rewards in game contexts (e.g., Allen & Lewis, 2019; Ding et al., 2014; Heyman et al., 2015; O'Connor & Evans, 2019; Zhao et al., 2021; see Lee & Evans, 2013 for a review). The limited research on academic cheating among younger children might be because in Western countries, where most of the research on academic cheating has been conducted, elementary school education has undergone reforms in the years since World War II that have de-emphasized performance goals (Krou et al., 2021). These reforms, which have included the reduction or elimination of exams, may have lessened concerns about the problem of academic cheating.

There are nevertheless important reasons to uncover the extent of academic cheating during the elementary school years, along with the factors that contribute to it. One reason is that the reforms in Western countries that de-emphasized performance goals did not extend to many other parts of the world (e.g., Asia, Africa, South America, and Eastern Europe). In these regions, exams remain one of the primary ways of evaluating student learning, starting in the elementary school years or even earlier. In some Asian countries (e.g., P.R. China), elementary school children take exams frequently (at least four major exams per year according to a central government mandate, typically with shorter tests between these exams as well). In addition, the elementary school years are a period during which attitudes about academic integrity are emerging but have not yet become entrenched. Given earlier evidence (Brandes, 1986; Hartshorne & May, 1928) and recent findings regarding academic cheating among 5- to 6-year-olds (Zhao et al., 2020), academic cheating may already be prevalent during the elementary school years.

Another major question the present study aimed to address is whether the key factors known to be significantly associated with academic cheating among high school and university students are also associated with elementary school children's cheating. Two possibilities exist. One is that elementary school children's cheating behavior may be affected by factors different from those affecting older students' cheating. This is because elementary school children have different levels of cognitive, social, moral, and neural development from high school and college students and they are exposed to substantially different micro- and meso-environments (Bronfenbrenner, 1993). If this *discontinuity* hypothesis is true, a new theoretical model may need to be developed to account for the emergence of academic cheating among younger students. Alternatively, it is also possible that the factors associated with academic cheating are largely the same for younger and older students because many of the characteristics of formal schooling remain constant across grade levels (e.g., regardless of the level of education and development, students typically learn along with peers of similar ages). If this *commonality hypothesis* is correct, the theoretical model proposed by Whitley (1998) may still be able to account for academic cheating in elementary school children.

Studying the emergence and early development of academic cheating also has important practical implications. It is now well established that by the end of high school, the number of students who self-report having engaged in academic cheating is as high as 80%–90% (see Murdock & Anderman, 2006). It is also the case that students who report cheating in high school are more likely to report cheating in college (e.g., Davy et al., 2007; Desalegn & Berhan, 2014). Further, despite century-long efforts, most cheating prevention and deterrence methods have failed to reduce cheating among high school and college students (Popoola et al., 2017; Volpe et al., 2008; Youmans, 2011; Zhao et al., 2021), perhaps due to the fact that cheating behavior is already normalized at this age. Such findings point to the importance of early intervention efforts. To be successful, these efforts will need to be based on a comprehensive understanding of academic cheating during the elementary school years, a time when cheating behaviors and associated beliefs are likely to be highly malleable (Azar & Applebaum, 2019; Hartshorne & May, 1928; Zhao et al., 2020). Such an understanding can provide useful guidance for developing early prevention and intervention programs to promote a culture of academic integrity (Ip et al., 2016).

The present study investigated academic cheating among a large sample of elementary school children (N>2000). We used Whitley's influential model and related findings as the basis for constructing a questionnaire, with self-reported academic cheating as the predicted variable. For the predicting variables, we asked children how common they considered cheating to be among their classmates, and how often they had directly observed a peer engaging in cheating. We asked these questions because Waltzer and Dahl (2022) found an important distinction between prevalence (commonality) and frequency, with the former referring to how widespread cheating is, and the latter referring to how frequently an individual engages in cheating. They pointed out that this distinction is akin to getting married, which typically is highly prevalent for a population but infrequent for an individual. Additional predicting variables included the extent to which children found cheating to be acceptable, the extent to which they thought their peers considered cheating to be acceptable, their perceptions of the effectiveness of various strategies that adults use to reduce cheating, their perceptions of how severe the consequences are for being caught cheating, as well as demographic information.

For our data analysis, instead of using the traditional statistical methods, we used the advanced machine learning approach to systematically investigate the key determinants that can explain and predict children's cheating on exams. In recent years, many psychological researchers have begun to apply various machine learning approaches to study psychological phenomena such as emotion (e.g., Just et al., 2017), deception (e.g., Bartlett et al., 2014), anxiety (Sun, Luo, et al., 2022), and psychopathology (e.g., Kessler et al., 2016; Livieris et al., 2018; Sun, Liu, et al., 2022). Recently, machine learning has also been used by developmental researchers to predict various outcomes in children (e.g., Bleidorn & Hopwood, 2018; Bruer et al., 2019; Zanette et al., 2016).

Modern machine learning approaches have several advantages over traditional statistical approaches. First, machine learning can help to improve the internal validity (Campbell, 1986; Diener et al., 2022) of its predictions. This is because traditional statistical approaches such as Generalized Linear Modeling (GLM) and Generalized Estimating Equations (GEE) typically put all of the data into the analysis at once. Because no data are set aside for validating the model, the result is one single model that is often overfitted. Such overfitting typically reduces the generalizability of the models to new datasets with new samples of participants. The modern machine learning approach overcomes this problem by randomly partitioning the data into training and testing subsets. The model fitting is first done on the training subset, and then tested on the testing subset, which is not used in training. The two subsets are then recombined and reshuffled to produce a new training subset and a new testing set to ensure that the model's performance is not influenced by any particular way of partitioning the data into training and testing subsets. This process is repeated for many times, which produces many models that predict the dependent variable with varying degrees of success.

How well the predictors are able to predict the dependent variable across the multiple data partitioning and recombination processes can be evaluated statistically, which essentially provides an assessment of the internal validity of the models' predictions (Campbell, 1986; Diener et al., 2022). Furthermore, by using the resultant distributions of model performances, one can estimate effect sizes and the likelihood of false positives (Type 1 errors).

Second, the modern machine learning approaches involve dividing the data into training, testing, and holdout subsets. Then the models from the training–testing process are further validated against the holdout subset, which has never been used in either training or testing. Thus, how well the models perform with this subset provides a statistical assessment of the generalizability of the models and their potential reproducibility with a new group of participants. In other words, this validation against the holdout subsets provides an assessment of the models' external validity (Campbell, 1986; Diener et al., 2022) and can help to address the current replication crisis in psychology and other disciplines (e.g., Blockeel & Vanschoren, 2007; Drubin, 2015).

Third, in the modern machine learning approaches, we can choose a variety of different machine learning algorithms to analyze data. They include multiple linear regression or logistic regression (LR) analyses, which were among the first types of machine learning algorithms to be used in science (Ng, 2017). Regression algorithms can be used to model the linear or curvilinear effects of predictors and their interactions. More modern algorithms include Random Forest (RF), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost) decision trees (Chan et al., 2002; Gao et al., 2018; Golino et al., 2014; Yarkoni & Westfall, 2017). These algorithms can model not only linear or curvilinear effects, but also dynamic nonlinear effects (Kurt et al., 2008; Stylianou et al., 2015; Thelen & Smith, 1994). Different algorithms produce different models with different levels of predictive performance, which allows researchers to determine the best technique for analyzing the dataset at hand. Furthermore, using multiple machine learning algorithms can help to reduce the odds of a failure to reveal a true significant association between the predicted variable and the predictors (a false negative or Type 2 error).

Fourth, modern machine learning approaches can use the Shapley values to explain their findings by quantifying the relative importance of different predicting variables (Ghorbani & Zou, 2019; Lundberg & Lee, 2017; Smith & Alvarez, 2021). Shapley values are based on game theory, and were proposed by a Nobel Prize winner Lloyd Shapley (1953). At the beginning, game theory was used only to solve the problem of distributing benefits in complex cooperative relationships. Its distribution principle is that the benefit obtained by individuals should be equal to the value of their contributions. Recently, some scholars have begun to use this approach to measure the relative contributions of all predicting variables in machine learning models, and it has become one of the important metrics for explaining machine learning results (Ghorbani & Zou, 2019; Lundberg & Lee, 2017; Smith & Alvarez, 2021; Sun, Liu, et al., 2022; Sun, Luo, et al., 2022). Because the Shapley values are additive mathematically, one can perform conventional statistical analyses on them to make probabilistic inferences about whether one predictor is significantly superior or inferior to another predictor in its contribution to a computational model's performance. In the present study, the Shapley values allowed us to identify factors that predict cheating, from the most important to the least important.

The present study tested the following hypotheses for both confirmatory and exploratory purposes. First, we hypothesized that the LR, RF, MLP, and XGBoost decision trees each would all be able to produce computational models that predict cheating significantly above the chance level. Second, we hypothesized that among these four different machine learning algorithms, the RF, MLP, and XGBoost would be able to produce computational models that predict cheating significantly better than the traditional LR. These hypotheses were tested for confirmatory purposes because they were based on the fact that existing theoretical models and related empirical evidence suggests that cheating is influenced by a multitude of moderating and mediating factors, and it is likely that some of these effects could be dynamically nonlinear. Thus, because the LR only considers the linear or curvilinear effects of the predictors on the predicted variable, this machine learning algorithm was predicted to perform worse than the other three algorithms, which are capable of considering linear, curvilinear, and nonlinear dynamic relations.

Third, regarding the predictor importance, we tested the *commonality* hypothesis for confirmatory purposes and hypothesized that, in line with this hypothesis, the important predictors identified by the computational models would align with the model proposed by Whitley et al. (1999). Specifically, in light of evidence of continuity between middle school and older students (Anderman & Midgley, 2004; McCabe et al., 2012), we predicted that the factors affecting cheating in elementary school students would be similar to those affecting cheating in older students. More specifically, we predicted that as with high school and college students, beliefs about the acceptability of peer cheating and observations regarding its frequency would be significantly associated with the likelihood of self-reporting to have cheated on exams (Krou et al., 2021; Lee et al., 2020; see Zhao et al., 2022 for related meta-analyses). Based on previous but relatively weak evidence among older students that males tend to cheat at a slightly higher rate than females (Jensen et al., 2002; Nathanson et al., 2006; Newstead et al., 1996; Whitley, 1998; Yu et al., 2016; see Whitley et al., 1999 for a meta-analysis), for exploratory purposes we hypothesized that this gender difference might already exist in elementary school children.

METHOD

Participants

The research was approved by the ethics committee of Hangzhou Normal University. Parents or legal guardians gave informed consent for their children to participate.

Children from second grade to sixth grade participated in the study. We did not include children in first grade because they have only limited experience with exams. The participants were recruited from three elementary schools in a city located in Hangzhou, a metropolitan city located in eastern China with a population of over 11 million. Of the 2300 surveys that were distributed, 2094 were considered valid, with the rest 206 excluded because they were less than 70% complete. Among these 2094 children, 555 were from school A where most students were from low middle SES families (average family income: 7144 yuan per month; education level: 8% college and above), 930 were from school B where most students were from middle SES backgrounds (average family income: 7856 yuan per month; education level: 58% college and above), and the remaining 609 were from school C where most students were from high SES backgrounds (average family income: 15,653 yuan per month; education level: 85% college and above).

The final sample (age: M = 10.05 years, SD = 1.40 years; 1111 boys) included 399 s graders (age: M = 7.88 years, SD = 0.63 years; 206 boys), 392 third graders (age: M = 8.78 years, SD = 0.40 years; 203 boys), 385 fourth graders (age: M = 9.79 years, SD = 0.33 years; 210 boys), 449 fifth graders (age: M = 10.74 years, SD = 0.33 years; 240 boys) and 469 sixth graders (age: M = 11.75 years, SD = 0.34 years; 252 boys). According to the school records, all participants were Han Chinese. Data collection was conducted in Mandarin and took place from October 15, 2018, to January 3, 2019.

Development of the survey

The process of creating the survey involved two design phases. In the first design phase, we conducted semi-structured interviews with nine teachers and 39 students from one of the schools where the research was conducted. Each interview was conducted in a one-on-one session that was developed based on Lim and See (2001). Each interview was audio recorded and coded by two research assistants who were blind to the purpose of the research. None of the students who contributed to either of the design phases participated in the final survey.

In the second design phase, a preliminary survey was created that was based on frequent responses from the first design phase, as well as Whitley's model and some related empirical studies among high school and college students (Bucciol et al., 2017; Lee et al., 2020; Özcan et al., 2019; Whitley, 1998; see Zhao et al., 2022 for a meta-analysis). The preliminary survey was then presented to a total of 158 s, fourth, and sixth graders.

Based on the results of the preliminary survey, we revised the items to produce a final survey, which was administered via in-person sessions at the children's school. The questionnaire was anonymous, and no teachers were present as children completed it. The surveys were distributed by researchers who gave instructions on how to complete it. The surveys contained the following sections. See the Appendix for the complete set of items.

Dependent measures

Self-reported cheating

We assessed self-reported cheating with the question, “Have you ever engaged in cheating during an exam?” Participants responded using a 5-point Likert scale that ranged from 1 (*never*) to 5 (*extremely frequently*). Because the ratings were heavily skewed toward the left (i.e., toward the *never* end of the scale), we recoded them into two yes/no measures to obtain the predicted variable *cheating on exams*, with those who selected *never* classified as not having engaged in cheating, and those who selected any other value classified as having engaged in cheating. This variable served as the predicted variable for the subsequent analyses.

Predictors

Observed peer cheating

This section consisted of three questions about the predictor variables.

Commonality (i.e., how widespread cheating is)

The first question was “To what degree do you think cheating on exams is widespread among your classmates?”. Responses ranged from 1 (*not widespread at all*) to 5 (*extremely widespread*).

General frequency

The second question was “How often do you think your classmates have engaged in cheating during an exam?” Responses ranged from 1 (*never*) to 5 (*extremely frequently*).

Specific frequency

The third question contained of a list of eight possible ways to cheat during an exam (Cronbach's α = .79). Participants were asked to rate each item in terms of how often they thought their classmates had engaged in such behavior (e.g., bringing unauthorized materials to an exam) on a 5-point scale ranging from 1 (*never*) to 5 (*extremely frequently*).

Acceptability of cheating

This section addressed the acceptability of cheating. Responses were made on a 5-point Likert scale ranging from 1 (*not acceptable at all*) to 5 (*completely acceptable*).

Acceptable to self

Participants were asked, "To what degree is cheating on exams acceptable to you?"

Acceptable to peers

Participants were asked, "To what degree do you think cheating on exams is acceptable among your classmates?"

Perceived effectiveness of strategies that adults use to reduce cheating

Participants were given a list of eight strategies that adults might use to limit cheating on exams (Cronbach's $\alpha = .714$). They were asked to rate each item (e.g., "working harder") in terms of its effectiveness in reducing cheating, using a 5-point Likert scale, ranging from 1 (*not effective at all*) to 5 (*extremely effective*). We will refer to these as the *effectiveness* measures.

Perceived consequences of being caught cheating

Participants were given a list of five possible consequences of academic cheating (e.g., being punished by one's teacher; Cronbach's $\alpha = .80$), and were asked to rate each item in terms of its severity ("To what degree do you think each of the following items is severe?") on a 5-point Likert scale that ranged from 1 (*not severe at all*) to 5 (*extremely severe*). We will refer to these as the *consequence* measures.

Demographic information

This section included questions about the following topics: *school*, *participants' age*, *gender*, *information about siblings* (categorical variable: the only child, having one or more older siblings, having one or more younger siblings, or having both older and younger siblings), and *achievement level* (i.e., "What level do you think your academic performance is in the class?").

Analytic approaches

We first used SPSS (version 25) to provide descriptive statistics about children's responses to all survey questions. Next, we used Scikit-learn: Machine Learning in Python (Pedregosa et al., 2011) to perform machine learning analyses that were both confirmatory and exploratory.

Because we had two categorical predictors with more than two categories, we dummy-coded these categorical variables. Specifically, school type had three categories and was therefore coded into two dummy variables (with school A as the reference). Information about siblings (i.e., the "only children or not" classification) had four categories and was coded into three dummy variables (with the only child as the reference).

Model training, testing, and validation

Some of the children did not complete the questionnaire. Because the machine learning algorithms we used require data with no missing values, and that artificially imputing missing values can induce unforeseen biases, we only used the data from children who responded to all of the questions. As a result, the total number of children whose data were entered into the machine learning analyses was 1339.

To conduct the machine learning algorithms, for each model we randomly divided children's data into three subsamples: 64% of the total sample for the training set, 16% for the testing set, and 20% for the holdout set. It should be noted that the sample size of the two categories of the predicted variable (i.e., 1 = *never* vs. all other responses) was heavily imbalanced (only 26% reported having cheated), which can affect the performance of machine learning algorithms. To address this issue, we used the Random Over Sampler method from the imblearn library in Python. This method is a commonly used strategy to achieve data class balancing (i.e., the sample size of cheating and no-cheating) that involves increasing the number of samples in the minority class by randomly sampling with replacement from the current available samples. After rebalancing the data, we divided the remaining children's data into the training set (64% of the total sample) and the test set (16% of the total sample).

We then proceeded to use the training set to train computational models using LR, MLP, XGBoost, and RF. LR is a traditional machine learning algorithm that uses a logistic function to model a categorical variable, commonly a binary variable as predicted by predictors in a linear combination. MLP neural network is a feedforward artificial neural network that consists of an input and output layer and at least one hidden layer with nodes connected with different weights. The XGBoost decision trees algorithm is a variant of the gradient boosting technique that combines decision trees with Stochastic Gradient Boosting with Regularized Gradient Boosting. RF is an ensemble learning algorithm that constructs a multitude of decision trees to make predictions about the predicted variable. Among the four machine learning algorithms, only LR assumes linearity between variables.

For MLP, the network consisted of a single hidden layer and a hyperbolic tangent (tanh) activation function. We normalized covariates prior to training. The training process used a scaled conjugate gradient descent algorithm (lambda = .0000005, sigma = .00005, interval center 0, interval offset ±.5) to adjust network connection weights in a way that minimized prediction error (for participants allocated to the training set) over successive training epochs. At the conclusion of each training epoch, the algorithm calculated prediction error in the testing set to ensure that error reductions achieved in the training set were not due to overfitting the training data. For XGBoost, in the model training process all parameters were set by default. In detail, tree-based models (gbtree) were used for booster. The number of gradient-boosted trees was set to 100. The Eta (boosting learning rate) was set to .3, which refers to set size shrinkage and is used in updates to prevent overfitting. All covariates were normalized for training and testing. For RF, in the training process, bootstrap samples were used when building trees. The number of trees in the forest was set to 100 by default and the number of features to consider when looking for the best split was set to the square root of the number of features. These two parameters were applied to improve the predictive accuracy and control over-fitting. All covariates were normalized for training and testing.

Next, we trained on the training subset and tested on the test subset to obtain the first model and its performance metrics. Then we combined the training and testing subsets and randomly re-partitioned the data into a new training subset and a new test subset to train the second model. We repeated the process 100 times to obtain 100 models to ensure that our findings could be reproduced irrespective of which subjects were partitioned into the training versus testing subset. After obtaining the 100 trained and tested models for each machine learning algorithm, we validated them against the pristine holdout set that was never used in training or testing to test the external validity of the model.

For both the testing and validation sets, the 100 models produced two key performance metrics. The first was accuracy, which refers to the extent to which children could be correctly classified by a model as having cheated or not cheated based on their responses to the predictor questions. The second was the area under the receiver operating characteristic curve. The area under curve (AUC; shown in Figure 1) is a common machine learning performance measure that incorporates a single index of a model's overall ability at classification with

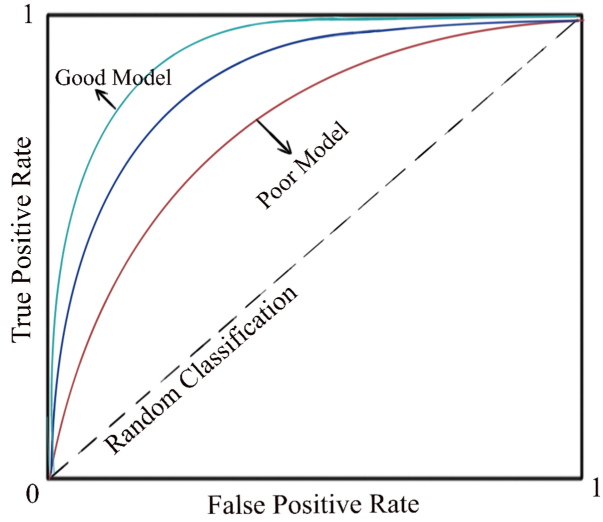

FIGURE 1 Receiver operating characteristic curve depicts how sensitivity (true positive) varies as a function of specificity (1 − false positive rate). The area under the curve depicts the overall performance of a model with the closer the curve to the line of identity (the diagonal dashed line) the poorer the model's ability to classify.

its sensitivity (i.e., how well the model predicts correctly students who have cheated as having cheated) and specificity (i.e., how well the model predicts children who have not cheated as not having cheated).

Based on the performance of the above two metrics, we selected the best technique to use for predicting whether individual children reported having cheated on exams. In addition, we computed the Shapley values of all predictors of all models based on this technique.

In recent years, researchers have begun to use the Shapley values to evaluate the relative contribution of each predictor in predictive models, and provide explanations about the findings based on machine learning (Ghorbani & Zou, 2019; Lundberg & Lee, 2017). The Shapley values were originally developed to deal with complex allocation problems (Shapley, 1953). To illustrate, imagine a situation in which three individuals A, B, and C complete a task together. When allocating bonuses to A, we need to obtain A's marginal contribution through the Shapley value method to ensure fairness. This approach involves calculating the amount of work that can be completed when only A is involved, the amount of work that can be completed when B and C cooperates with A minus the amount of work that can be completed by B and C alone, the amount of work that can be completed when B and C cooperate with A minus the amount of work that can be completed when B only cooperates with C, and so on. Computing the average of these values produces the final marginal contribution of A. In sum, Shapley values comprehensively consider all possible situations and perform a fair calculation of the marginal contribution of each entity in a principled manner.

To relate this example to the present study, A, B, and C refer to the different predictors, and the amount of work refers to the contribution of the predictors in determining the accuracy of model's prediction. In other words, the Shapley value of our predictors represents its actual marginal contribution to the prediction accuracy of the model. Thus, by calculating the Shapley values, we can statistically assess which of the variables are the most important predictors of children's cheating and which are relatively less important.

RESULTS

Dependent measures

Self-reported cheating

Overall, 25.74% of children reported cheating on exams. Figure 2 shows a violin plot of the distributions of children who self-reported having cheated or not as a function of age. We performed a Pearson correlation analysis using children's age in years as a continuous variable and found that self-reported cheating did not change significantly with age ($r = .03$, $p = .204$).

Responses to predictor questions concerning cheating on exams

Table 1 shows the average ratings for each predictor of cheating on exams. As can be seen from the table, children tended to view cheating negatively ($M = 1.64$ on a 5-point scale, with 1 being *not acceptable at all*), and they believed that their peers view it slightly less negatively than they do ($M = 1.91$ on the same scale; difference in ratings $p < .001$ by a paired-samples t-test). Moreover, "copying answers from a neighbor during an exam" was rated as the form of cheating on exams that was engaged in most frequently ($M = 1.96$), "working harder to pass the exam" was rated as the most effective way to reduce cheating on exams ($M = 3.57$), and "being punished by teachers" ($M = 3.65$) was rated as the most severe consequence of being caught cheating.

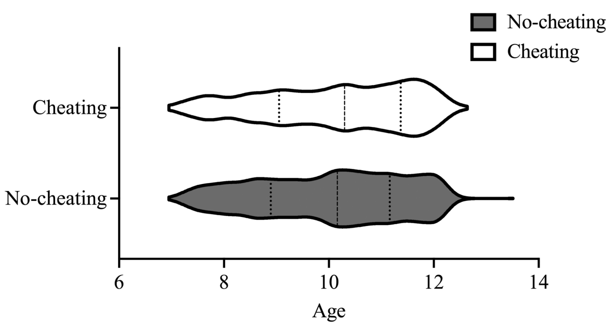

FIGURE 2 Violin plot of self-reported cheating on exams by age in years.

TABLE 1 Descriptive results for predictor questions about cheating on exams.

Item	*M*	SD
Q2. Commonality (How widespread?)	1.87	0.81
Q3. General frequency (How often?)	1.91	0.87
Q4. Specific frequency, 1–8 (How often?)		
1. Bringing unauthorized materials to an exam	1.42	0.70
2. Copying answers from a textbook during an exam	1.39	0.65
3. Passing notes during an exam	1.32	0.66
4. Copying answers from a neighbor during an exam	1.96	0.99
5. Using tools (e.g., dictionaries, cellphones, smart watches) to search for answers during an exam	1.13	0.45
6. Working with one or more classmates to share answers during an exam	1.63	0.89
7. Deliberately giving a classmate a wrong answer during an exam	1.33	0.72
8. Secretly changing a test score	1.09	0.39
Q5. Acceptable to self	1.64	1.05
Q6. Acceptable to peers	1.91	1.11
Q7. Effectiveness, 1–8		
1. Increasing the severity of consequences for getting caught cheating (e.g., giving a zero score)	3.21	1.44
2. Students who sit next to each other getting different versions of the test	2.93	1.54
3. Working harder	3.57	1.49
4. Teachers emphasizing that academic cheating represents a serious moral transgression	2.62	1.38
5. Having a teacher who is greatly liked by their students teach the class	2.07	1.30
6. Teachers giving sharp criticism or punishment	3.09	1.32
7. Teachers recognize classroom role models by giving praise or prizes to students who behave honestly on exams	3.00	1.40
8. Parents being informed when their children are caught cheating, and the parents in turn giving sharp criticism or punishment	3.05	1.37
Q8. Consequence, 1–5		
1. Being criticized by one's teachers	3.05	1.22
2. Being punished by one's teachers	3.65	1.31
3. Being criticized by one's parents	3.24	1.25
4. Being punished by one's parents	3.36	1.30
5. Being criticized or rejected by one's classmates	3.33	1.48

Machine learning results

Accuracy

We found that all machine learning algorithms were able to predict children's cheating on exams significantly above the chance level (50%, $ps<.001$) for both the testing set (Table 2) and the holdout set (Table 3). Second, we found that the tree-based RF algorithm, performed significantly better than the traditional LR for both the testing set and holdout set ($ps<.05$). Third, the mean accuracy of the RF algorithm was 78.56% for the testing set and 81.43% for the holdout set, which was significantly better than the other machine learning algorithms ($ps<.05$). Fourth, for the testing set, MLP and XGBoost significantly outperformed the traditional LR. For the holdout set, XGBoost did not significantly outperform the traditional LR. Thus, the use of more modern machine learning algorithms did not always guarantee higher performance relative to the traditional LR algorithm.

AUC

We found that the AUC was significant above the chance level for all machine learning algorithms (50%, $ps<.001$), for both the testing set (Table 4) and the holdout set (Table 5). For the testing set, none of the three algorithms (XGBoost, 75.18%; MLP, 74.43%; RF, 74.79%) significantly outperformed the traditional LR (74.37%).

TABLE 2 Means and SDs of the accuracies (%) of the 100 models for predicting self-reported cheating on exams when tested on the testing set for the four machine learning algorithms.

Model	*M*	SD	95% Confidence interval for mean	
			Lower	Upper
Logistic Regression	69.71	2.97	69.12	70.30
XGBoost	71.53	3.22	70.89	72.17
Multilayer perceptron	77.44	2.19	77.01	77.88
Random Forest	78.56	2.41	78.08	79.04

TABLE 3 Means and SDs of the accuracies (%) of the 100 models for predicting self-reported cheating on exams when validated on the holdout set for the four machine learning algorithms.

Model	*M*	SD	95% Confidence interval for mean	
			Lower	Upper
Logistic Regression	75.48	1.10	75.26	75.70
XGBoost	75.70	1.82	75.34	76.06
Multilayer perceptron	80.19	0.74	80.04	80.34
Random Forest	81.43	0.86	81.25	81.60

TABLE 4 Means and SDs of the AUC (%) of the 100 models for predicting self-reported cheating on exams when validated on the testing set for the four machine learning algorithms.

Model	*M*	SD	95% Confidence interval for mean	
			Lower	Upper
Logistic Regression	74.37	3.51	73.68	75.07
XGBoost	75.18	3.58	74.47	75.89
Multilayer perceptron	74.43	3.14	73.82	75.05
Random Forest	74.79	3.23	74.15	75.43

TABLE 5 Means and SDs of the AUC (%) of the 100 models for predicting self-reported cheating on exams when validated on the holdout set for the four machine learning algorithms.

Model	*M*	SD	95% Confidence interval for mean	
			Lower	Upper
Logistic Regression	77.66	0.77	77.51	77.81
XGBoost	80.02	0.95	79.83	80.21
Multilayer perceptron	78.16	0.74	78.01	78.31
Random Forest	80.29	1.12	80.07	80.51

However, for the holdout set, the mean AUC of the RF algorithm and XGBoost algorithm reached 80.29% and 80.02%, respectively, which were significantly better ($ps<.05$) than that of the two other machine learning algorithms ($ps<.05$). Further, for the holdout set, MLP did not significantly outperform traditional LR, which reinforces the point that the more modern machine learning techniques are not always better than the traditional LR that only considers linear and curvilinear combinations. However, RF algorithms performed the best and produced models that on average had an 80.29% likelihood of correctly classifying whether children in the holdout set had cheated, with high specificity and sensitivity.

Because the performance of the RF models was the best in terms of classification accuracy, in Figure 3 we present only the AUC of the 100 models based on this algorithm when validated against the holdout set. As can be seen from the figure, the overall model had both high sensitivity and specificity (1 − false positive rate).

In theory, one could tune the hyperparameters of the RF models to improve their performance further. However, when we tried to tune the model hyperparameters of the best models based on RF, we found that the AUCs were not significantly better than the original ones. It appears that given the data at hand, we have hit the performance ceiling. Had we collected data from more children, more information from them (e.g., at multiple time points), or from additional informants (e.g., teachers and parents), we might have been able to obtain models with better performance.

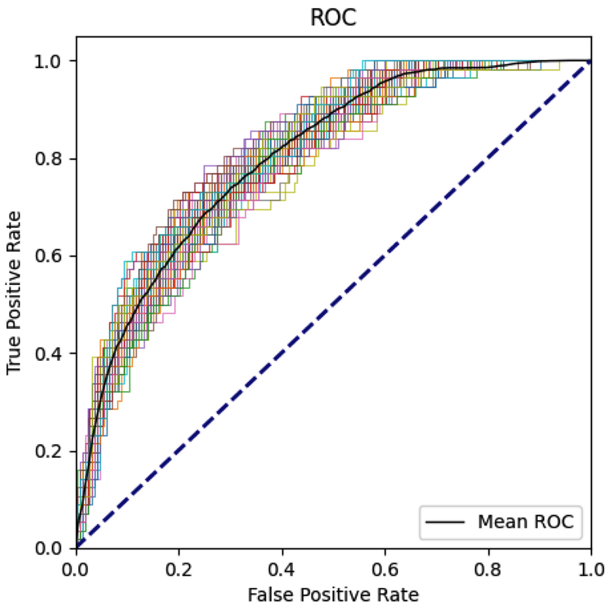

FIGURE 3 Mean area under the curve and 95% confidence intervals of the 100 models based on Random Forest against the holdout set. ROC, receiver operating characteristic.

It should be also noted that our LR algorithm does not consider higher order interactions, whereas the rest of the algorithms automatically do by default. Thus, the comparisons between the performances of the logistic models and the other models seemed to be unfair. However, the more modern algorithms such as RF, MLP, and XGBoost can automatically identify linear and nonlinear combinations of predictors to predict the outcome variable with optimized computational demands. In contrast, the LR algorithm requires handcrafted interactive terms, which is manageable when the number of the predictors is small. However, when the number of predictors is large, as in the present study (33 predictors), modeling all possible interactions would have unacceptably high computation demands, and also reduce the statistical power to produce robust and generalizable models. Nevertheless, to be certain, we performed another set of machine learning analyses using the LR algorithm that included all the predictors as the first-order variables and all possible two-way interactions between the categorical predictors and each of the continuous predictors. We found that the accuracies for training, testing, and holdout validation were 79.2%, 66.7%, and 67.6% respectively, and the AUCs for training, testing, and holdout validation were 88.1%, 68.9%, and 69.1%, respectively. Clearly, these models trained even with only a minimal number of interactive terms overfitted and, as a result, they were poorly generalizable to either the testing or the holdout dataset. Further, the accuracies and AUCs for the testing and holdout sets were significantly poorer than their counterparts when we used the logistic models without the interactive terms ($ps < .01$; see Tables 2–5).

Shapley values

Because the performance of the RF model was the best in terms of classification accuracy, we only reported the Shapley values of all predictors across 100 RF model iterations when validated on the holdout set. We found all values to be significantly above zero ($ps < .05$), which means all predictors made significant marginal contributions to the ability of the RF models to predict whether students had reported to having cheated on exams.

We arranged these predictors according to the size of the Shapley value to visually display the relative importance of these predictors in the RF models in Figure 4 by showing only the main predictors (and their 95% confidence intervals) whose marginal contribution rate to the accuracy of the models' prediction as greater than 1%. In Table 6, we show the means, SDs, and 95% confidence intervals of the Shapley values for the rest of the predictors that had values less than 1%.

It should be noted that when interpreting the Shapley's importance value for each predictor's ability to predict an outcome variable one must qualify it in relative terms, because each value is computed when the contributions from all predictors are considered together. Thus, relatively speaking, the predictors shown in Figure 4 were significantly more important than those in Table 6 when the contributions of all the predictors were considered together.

The Shapley values of the more important predictors in Figure 4 varied considerably, and they formed roughly four groups. The first group is whether the children considered cheating to be acceptable (with the contributions of all other predictors considered together, the more acceptable children found cheating to be, the more likely they were to self-report having cheated on exams). The Shapley value of this predictor reached 7.07%, which represents an actual marginal contribution to the models' prediction accuracy of 7.07%. The paired-sample t-test results show that the Shapley value of this variable is significantly greater than 4.20% (the Shapley value of the second-ranked predictor variable, $t = 18.65$, $df = 268$, $p < .001$). This result indicates that the children's view of the acceptability of cheating was a significantly better predictor than any of the other measures.

Within the second most important group of predictors, the most important one was how frequently children had observed peers working with one or more classmates to share answers (4.20%; with the contributions of all other predictors considered together, the more they had observed such events, the more likely they were to self-report having cheated). The next predictors were how widely children believed cheating to be occurring at their school (3.18%; with the contributions of all other predictors considered together, the more widespread children perceived peer cheating to be, the more likely they were to self-report having cheated) and how frequently the children had observed peers copying answers from a

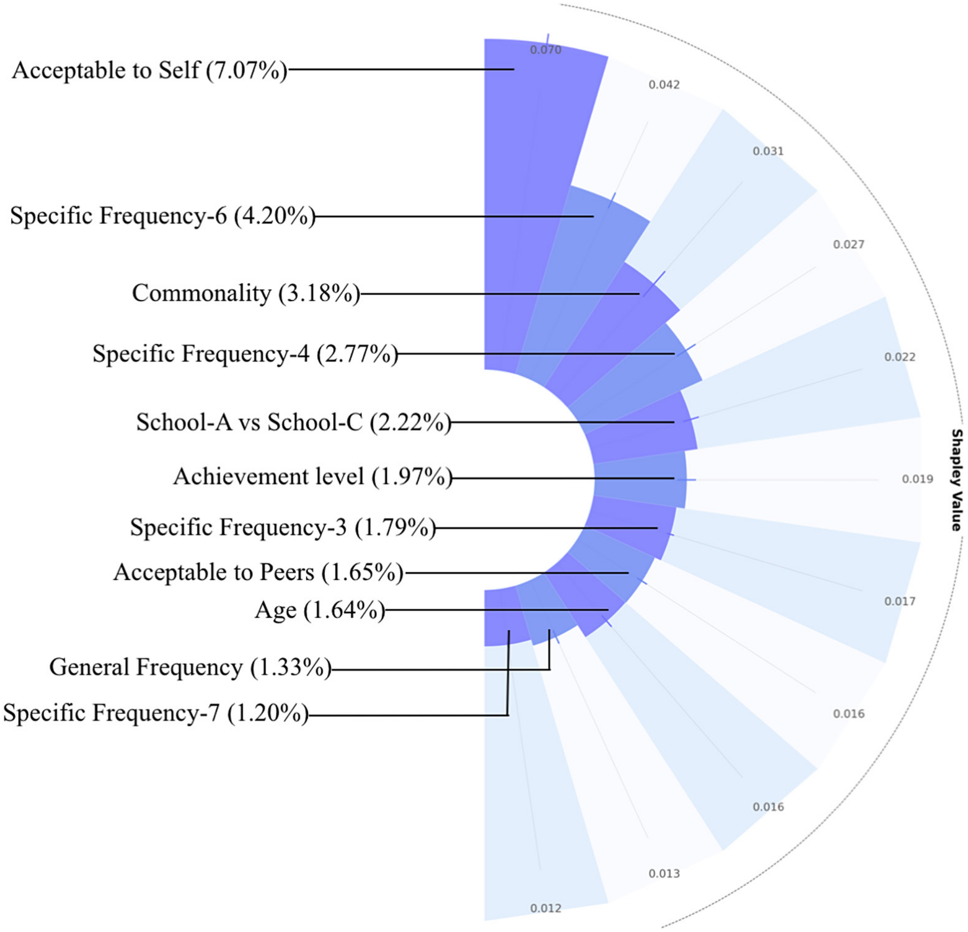

FIGURE 4 Mean predictor Shapley importance values and 95% confidence intervals across all 100 Random Forest models when validated on the holdout set to predict self-reported cheating on exams (information on the main predictors with Shapley values higher than 1%).

neighbor during an exam (2.77%; with the contributions of all other predictors considered together, the more frequently children observed such behavior, the more likely they were to self-report having cheated). The Shapley values of these predictors were significantly higher than that of all remaining variables (4.20%, 3.18%, and 2.77% compared to 2.22%, t = 14.45, 4.68, and 3.53, dfs = 268, ps<.001).

The third important group of predictors included the contrast between School A (the lowest SES) and School C (the highest SES; 2.22%; with all other predictors held equal, students from School A were more likely to self-report having cheated than those from School C), children's own academic achievement level (1.97%; with the contributions of all other predictors considered together, children who reported a lower level of achievement were more likely to self-report having cheated).

The Shapley values of the fourth important group were all in the range from 1% to 2%, which means their marginal contribution to the model was very small. These predictors include: whether cheating is considered acceptable by peers in general (1.65%; with the contributions of all other predictors considered together, the more acceptable they thought cheating is to peers, the more likely they were to self-report having cheated), age (1.64%; with the contributions of all other predictors considered together, older children were more likely to self-report having cheated), the overall frequency of observing peers cheating (1.33%; with the contributions of all other predictors considered together, the higher the frequency of observed peer cheating, the more likely children were to self-report having cheated), how frequently they have observed peers passing notes and deliberately giving a classmate a wrong answer during an exam and deliberately giving a classmate a wrong answer during an exam (1.79% and 1.20%; with the contributions of all other predictors considered together, the more frequently children had observed it, the more likely they were to self-report having cheated).

In addition to the above four groups of predictors, the predictors that are shown in Table 6 but not in Figure 4 had very weak importance values (Shapley values <1%), although they were all significantly above zero (ps<.05). These predictors included children's beliefs about the effectiveness of different cheating deterrence strategies and children's beliefs about the severity of different cheating

TABLE 6 Mean predictor Shapley importance values (%), standard deviations, and 95% confidence interval across all 100 model iterations when validated on the holdout set to predict self-reported cheating on exams (information on the minor predictors with Shapley values lower than 1%).

			95% CI	
Item	***M***	**SD**	**Lower**	**Upper**
Effectiveness–8	0.99	0.49	0.93	1.05
Effectiveness–7	0.95	0.86	0.85	1.06
Effectiveness–5	0.88	0.60	0.80	0.95
Effectiveness–4	0.79	0.45	0.74	0.85
Consequence–5	0.76	0.52	0.70	0.83
Specific frequency–2	0.71	0.25	0.68	0.74
"Have one or more older siblings" versus "The only child"	0.70	0.30	0.66	0.73
Effectiveness–3	0.68	0.50	0.62	0.74
Effectiveness–2	0.68	0.51	0.62	0.74
Consequence–4	0.66	0.36	0.62	0.71
Consequence–3	0.64	0.51	0.57	0.70
Effectiveness–6	0.63	0.51	0.57	0.69
Consequence–1	0.62	0.57	0.55	0.69
Consequence–2	0.60	0.47	0.54	0.66
Effectiveness–1	0.59	0.43	0.54	0.64
Specific frequency–1	0.47	0.29	0.44	0.50
"Have one or more younger siblings" versus "The only child"	0.34	0.17	0.32	0.36
Gender (girl vs. boy)	0.32	0.25	0.29	0.35
Specific frequency–5	0.29	0.16	0.27	0.31
School (School A vs. School B)	0.19	0.14	0.18	0.21
"Have both older and younger" versus "The only child"	0.18	0.51	0.12	0.24
Specific frequency–8	0.13	0.39	0.08	0.18

Note: See Table 1 and Appendix for the exact wording of the questions.

consequences (Table 6). Table 6 shows that boys were significantly more likely to report cheating than were girls, which is consistent with previous findings among older students (see Cizek, 1999 for a review; Whitley et al., 1999 for a meta-analysis). However, this finding should be interpreted with caution because the Shapley value, though significant, was very small (.3%). The same is true about the contribution of having a sibling, whose Shapley value was significant but very small (≤.7%; children with siblings were more likely to self-report having cheated).

DISCUSSION

The present study used a machine learning approach to examine elementary school children's self-reported cheating in relation to a range of factors that, according to previous research, are linked to self-reported academic cheating in older students. We obtained several major findings. First, we found that overall about 25.74% of children reported to have cheated on exams. This level of self-reported cheating is on the low end of the 20%–94% range that has been reported in studies involving older students (Desalegn & Berhan, 2014; Hrabak et al., 2004). The finding supports our hypothesis that self-reported academic cheating emerges during the elementary school years, if not earlier. However, as compared to the high rate of cheating among older students, the cheating rate in elementary school children was relatively low and it did not increase with age, which is comparable to the few previous studies involving elementary school children (Alan et al., 2019; Hartshorne & May, 1928). Taken together, previous research and the present findings suggest that self-reported cheating rates may be relatively low during the elementary school years, with a rapid increase during the middle school and high school years (Brandes, 1986; see Jensen et al., 2002).

The relatively low self-reported academic cheating rate among elementary school children also stands in contrast with rates that have been reported by researchers who have examined cheating in games. Such cheating begins as early as 2 years of age, and cheating rates can reach as high as 70%–90% (e.g., Allen & Lewis, 2019; Ding et al., 2014; Heyman et al., 2015; O'Connor & Evans, 2019; Zhao et al., 2021; see Lee & Evans, 2013 for a review). One reason may be that during the elementary school years the academic rewards children can gain by cheating at school are less tangible than the material rewards they can gain by cheating in games (see Lee, 2013). Another reason may be that self-report measures underestimate the actual prevalence of cheating (Simpson & Yu, 2012), a possibility that awaits future empirical evaluation.

Second, machine learning results based on the RF with the holdout data show that the mean prediction accuracy of the computational models was relatively high (81.43%), which indicates that it was able to correctly predict cheating more than 80% of the time, based on the predictor variables. In addition, the mean AUC of our computational model was 80.29% for the holdout set, suggesting that this model was not only accurate, but also had reasonable sensitivity and specificity. The converted Cohen's d based on the AUC results suggests that the final machine learning model produced large effect sizes (Cohen's $d > 1.2$; Cohen, 1988) when using the predictor variables to predict children's cheating. Both the prediction accuracy and AUC of this algorithm were significantly higher than those of the traditional LR algorithm, which supports our hypothesis.

Our results also demonstrate that the more modern machine learning algorithms are not always superior to logistic models. In fact, the XGBoost models were not significantly better than the LR models in terms of accuracy. Moreover, the MLP models were also not significantly better than the LR models in terms of AUC. Thus, whether a specific machine learning algorithm is specifically suited to predict academic cheating is an empirical question that needs to be tested with each dataset. For the present dataset, RF is clearly the best among the four algorithms explored here. Nevertheless, for other datasets, LR or other machine learning algorithms may be better than RF (see Abaker & Saeed, 2021; Sufriyana et al., 2020 for examples).

Third, to address which predictors contributed the most to our accurate prediction of children's self-reported cheating, we computed the Shapley values for all predictors in the RF models. We found that although all predictors significantly contributed to the RF algorithm's high performance, their importance values varied significantly. First of all, when the contributions of all predictor variables were considered together, children's view of the acceptability of cheating was the single most important predictor of cheating by a large margin: the more acceptable children considered cheating to be, the more likely they were to self-report having cheated. This is in line with findings from older students (Abaraogu et al., 2016; Ives & Giukin, 2020) and it is consistent with Whitley's model (1998). Murdock and Anderman (2006) proposed that students consider the acceptability of cheating based on two factors: their moral beliefs about cheating (the more negatively they judge cheating, the less acceptable it is), and whether they are able to provide reasonable justifications for cheating (the more they can justify cheating, the more acceptable they judge it to be). Further, they suggested that these two factors work together to shift students' perception of behavioral norms regarding cheating on exams (i.e., what is morally acceptable) as well as reducing the impact of cheating on positive self-perception. Because this mechanism suggested by Murdock and Anderman (2006) is based on data from older students, it remains unclear whether it also applies to elementary school children, and this is an issue that should be explored in future studies.

The second most important predictor was how frequently children observed their peers to have cheated on exams, and this results is highly consistent with Whitley's model, as well as with existing findings from studies involving older students (McCabe & Abdallah, 2008; Rettinger & Kramer, 2009; see Zhao et al., 2022 for a meta-analysis). These predictors include observations of peers cheating in general, as well as observations of specific forms of cheating, such as sharing answers with classmates, or copying answers from a neighbor while taking exam. Students who scored highly on these measures were more likely to report having cheated themselves. Children's observations of how widespread cheating is at their school (i.e., its commonality) was the third most important predictor, which is consistent with previous findings involving older students, and is again in line with Whitley's model (1998).

These findings regarding commonality and frequency of cheating point to the importance of peer influences in this domain (Ghanem & Mozahem, 2019; Meiseberg & Ehrmann, 2016; Rettinger & Kramer, 2009). Previous studies with older students have found a robust "perceived peer cheating effect" whereby the perception of cheating by peers is positively correlated with children's own academic cheating (Ghanem & Mozahem, 2019; McCabe & Treviño, 1993, 1997; Meiseberg & Ehrmann, 2016; Whitley, 1998). In fact, several narrative reviews (Cizek, 1999; Murdock & Anderman, 2006) and a meta-analysis (Zhao et al., 2022) have concluded that perceived peer cheating is one of the most important aspects of academic dishonesty for students starting in middle school. The present findings suggest that this conclusion applies to elementary school children as well.

Several theories have been proposed to explain the influence of peer behavior on academic cheating. One theoretical approach is based on social learning theory (Bandura, 1969, 1977), and it suggests that a person who witnesses socially significant individuals engaging in and benefiting from a particular behavior is more likely to engage in similar behaviors themselves, even if the behavior violates societal norms (O'Rourke et al., 2010). Another approach, neutralization theory, suggests that individuals devise strategies that allow them to justify the social norms they violate, so that they can maintain a positive self-image (e.g., Pulvers & Diekhoff, 1999; Rettinger & Kramer, 2009). One neutralizing technique common in cheating contexts involves the claim that "everyone else is doing it" (e.g., Haines et al., 1986). This technique reduces or displaces one's own responsibility by attributing the causes of behavior to others, or to external factors (Stephens, 2017). A social-cognitive theory suggests that students who perceive cheating to be common may believe they are at a competitive disadvantage if they do not cheat, and may believe that the consequences of cheating are low if many students are able to cheat without facing consequences for it. The present research suggests that these theories may be applicable to cheating among elementary school children as well. However, additional empirical studies are needed to test the validity of these theoretical accounts with elementary school children because they are clearly at a different developmental period from older students, for whom peer relationships are known to play a much greater role in their socialization (Inderbitzen, 1994; Kafle & Thakali, 2013).

The other important predictor was school type. This may be related to the SES level of the school that children attended. Those from the lowest SES school (School A) were more likely to self-report having cheated than those from the highest SES school (School C). This finding is consistent with historical and recent studies of older students suggesting a higher rate of cheating among students from low SES families and neighborhoods (e.g., Alt, 2015; Hartshorne & May, 1928; Yu et al., 2016). This school effect might reflect a number of issues, including the school leadership, teacher cooperation and consensus, and school ethos (e.g., Ramberg & Modin, 2019; see McCabe et al., 2012 for a review). In addition, children's own academic achievement level was an important predictor too. Also in line with previous evidence among high school and university students (Huang et al., 2015; Koscielniak & Bojanowska, 2019), children who reported a lower level of academic achievement were more likely to report that they cheated on exams. This may be because children with poor academic performance feel a need to resort to cheating to avoid the negative consequences of failure (see Oran et al., 2016). In addition, students with high achievement levels might be more driven by the desire to learn than to earn good grades (Putarek & Pavlin-Bernardic, 2020).

Finally, we found that the Shapley values for children's beliefs about the consequences of cheating, and for their beliefs about the effectiveness of cheating deterrence strategies, as well as demographic characteristics (e.g., age and gender) were all significantly above zero. However, their importance values are small or very small, indicating that they were among the least important predictors of children's cheating. These results are not surprising, given the weak findings in the literature regarding these predictors. For example, some previous studies have found that male students are more likely to cheat than female students are (Jensen et al., 2002), whereas other studies have found no significant gender difference (Nathanson et al., 2006; see Whitley et al., 1999 for a meta-analysis). Similarly, there are inconsistent results in studies examining age associations. Some studies have found no significant association between age and academic cheating (Ives et al., 2017; Lambert et al., 2003), while others have shown different forms of association: Some have found that older students cheated more than younger students (Błachnio, 2019; Michaels & Miethe, 1989), whereas other studies have found the reverse pattern (Azar & Applebaum, 2019; Marsden et al., 2005; Stiles et al., 2018). In addition, some studies have found that cheating and perceived negative consequences are positively correlated (McCabe & Treviño, 1993), whereas others have found the opposite pattern (McCabe & Treviño, 1997).

In sum, the present findings reveal that the major factors known to be related to older students' academic cheating are also significantly associated with self-reported cheating among elementary school children (Abaraogu et al., 2016; Ghanem & Mozahem, 2019). Although additional data are still needed, the present findings nevertheless suggest that the theoretical model developed for older students (Whitley, 1998) does a good job of accounting for self-reported academic cheating in the elementary school years. This is particularly true with regard to attitudes toward cheating. In other words, the present findings support the *commonality hypothesis* more than the *discontinuity hypothesis*. Thus, there may be similar mechanisms underlying academic cheating in elementary school children and among older students.

In addition to the theoretical contributions, the present findings have practical implications. First, given the present finding that children's beliefs about the acceptability of cheating are the most important predictor of cheating, pedagogical efforts may need to focus on educating children about the value of academic integrity and persuading them to reject the view that cheating is acceptable. Second, given the importance of children's observation of the commonality and frequency of peer cheating in predicting their own cheating, we should strive to create a school environment where children are surrounded with peers who practice academic integrity so as to reduce negative peer influences. Third, we should strive to create a learning atmosphere where exams are used as a means to assess how individual children are progressing, rather than as a method of ranking students based on their academic performance (Cochran, 2015). Finally, the elementary school years may be an optimal time period to begin promoting academic integrity before cheating becomes highly prevalent and normalized given that cheating begins to increase rapidly during the middle school and high school years (Brandes, 1986; see Jensen et al., 2002).

The methodological contributions of the present study should also be noted. Over the last decade, there have been rapid advances in machine learning. Many advanced techniques have become available that consider relations among variables that the traditional GLM and GEE approaches are incapable of dealing with. As a result, some of these methods are able to produce predictive models with significantly higher performance than the traditional approaches. In addition, the data partitioning strategy that divides data into training, testing, and holdout subsets can lead to predictive models that are robust, generalizable, and with high external validity (Campbell, 1986). The present study, along with several previous ones (e.g., Bruer et al., 2019; Zanette et al., 2016), illustrates the viability of using machine learning approaches to analyze developmental data from children. Recent research in other areas of psychology (e.g., clinical psychology, social psychology, neuropsychology) have used machine learning to uncover previously unknown effects and provide new insights about various psychological phenomena (e.g., Bartlett et al., 2014; Just et al., 2017). For example, Just et al. (2017) were able to use machine learning to identify people with high suicide intentions with high accuracy (up to 91%) based on neural responses to specific words such as death, trouble cruelty, carefree, good, and praise. Bartlett et al. (2014) used machine learning to analyze the facial expressions of individuals who experienced real pain versus faked pain. They discovered that while the two groups of individuals displayed the same kind of facial expressions, using the dynamics of the expressions led to the accurate detection of pain faking (93% accuracy). We hope that the success of studies that use machine learning will provide an impetus for more developmental researchers to use these techniques to address a range of questions and advance our knowledge about child development.

The present study also has several limitations. One is that it only involved children from China. Although our findings are similar to what has been found with older students in the West, they need to be replicated with elementary school children in other countries. Second, in the present study, cheating was assessed using self-report measures only. Even though we took great care to reassure students that their responses would be kept confidential, some children might not have responded truthfully out of a fear that their responses would be accessible to their teachers. Thus, the findings of the present study may have underestimated of the actual prevalence of cheating and the strength of its relations to the predictor variables. Future studies should use self-report and behavioral cheating measures together to address this problem. Since Hartshorne and May (1928), researchers have devised many ingenious and naturalistic methods to assess whether children have cheated on a test. Although most of these methods have been used with older children and adults (Cizek, 1999; Zhao et al., 2021), recent studies have shown that they can be readily used with young children as well (Zhao et al., 2018, 2020, 2021). Another limitation of the present research is that we only examined correlations between variables at one time point. This leaves open questions about how beliefs and behavior influence each other and unfold over developmental time. This limitation could be overcome by using longitudinal designs.

CONCLUSIONS

The present study used a machine learning approach to examine self-reported academic cheating among Chinese second to sixth graders, to bridge a significant gap in the literature regarding the emergence of cheating during the elementary school years. We found that children's cheating was most strongly predicted by their own beliefs about how acceptable it is, by their observations of how prevalent cheating is at their school, and by how frequently they observe peers cheating. These factors are also significantly associated with academic cheating in older students, favoring the commonality hypothesis over the discontinuity hypothesis. The present study also provides important insights into how to promote academic integrity and limit cheating before it becomes entrenched. More broadly, the present research shows that machine learning is a viable and effective approach to analyzing developmental data.

ACKNOWLEDGMENTS

The contribution of L. Zhao was supported by the Major Project of Philosophy and Social Science Research of the Ministry of Education of China (22JZD044) and the National Natural Science Foundation of China (32171060). The contribution of Y. Zheng was supported by the Scientific Research Fund of Zhejiang Provincial Education Department (Y202250508). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DATA AVAILABILITY STATEMENT

The data and the code that support the findings of this study are available from L. Zhao (email: zhaoli@hznu.edu.cn) upon request.

ORCID

Li Zhao https://orcid.org/0000-0003-3286-4041
Gail D. Heyman https://orcid.org/0000-0001-7764-3205
Kang Lee https://orcid.org/0000-0001-5695-2615

REFERENCES

Abaker, A. A., & Saeed, F. A. (2021). A comparative analysis of machine learning algorithms to build a predictive model for detecting diabetes complications. *Informatica*, *45*, 117–125. https://doi.org/10.31449/inf.v45i1.3111

Abaraogu, U. O., Henning, M. A., Okpare, M. C., & Rajput, V. (2016). Disclosing academic dishonesty: Perspectives from Nigerian and New Zealand health professional students. *Ethics & Behavior*, *26*, 431–447. https://doi.org/10.1080/10508422.2015.1055494

Alan, S., Ertac, S., & Gümren, M. (2019). Cheating and incentives in a performance context: Evidence from a field experiment on children. *Journal of Economic Behavior & Organization*, *179*, 681–701. https://doi.org/10.1016/j.jebo.2019.03.015

Allen, J. W. P., & Lewis, M. (2019). Who peeks: Cognitive, emotional, behavioral, socialization, and child correlates of preschoolers' resistance to temptation. *European Journal of Developmental Psychology*, *17*, 1–23. https://doi.org/10.1080/17405629.2019.1665014

Alt, D. (2015). Assessing the connection between self–efficacy for learning and justifying academic cheating in higher education learning environments. *Journal of Academic Ethics*, *13*, 1–14. https://doi.org/10.1007/s10805–015–9227–5

Anderman, E. M., & Midgley, C. (2004). Changes in self–reported academic cheating across the transition from middle school to high school. *Contemporary Educational Psychology*, *29*, 499–517. https://doi.org/10.1016/j.cedpsych.2004.02.002

Azar, O. H., & Applebaum, M. (2019). Do children cheat to be honored? A natural experiment on dishonesty in a math competition. *Journal of Economic Behavior & Organization*, *169*, 143–157. https://doi.org/10.1016/j.jebo.2019.11.007

Bandura, A. (1969). Social-learning theory of identificatory processes. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (Vol. *213*, p. 262). Rand McNally College Publishing Company.

Bandura, A. (1977). *Social learning theory*. General Learning Press.

Bartlett, M. S., Littlewort, G., Frank, M., & Kang, L. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology*, *24*, 738–743. https://doi.org/10.1016/j.cub.2014.02.009

Błachnio, A. (2019). Don't cheat, be happy. Self–control, self–beliefs, and satisfaction with life in academic honesty: A cross–sectional study in Poland. *Scandinavian Journal of Psychology*, *60*, 261–266. https://doi.org/10.1111/sjop.12534

Bleidorn, W., & Hopwood, C. J. (2018). Using machine learning to advance personality assessment and theory. *Personality and*

Social Psychology Review, *23*, 108886831877299. https://doi.org/10.1177/1088868318772990
Blockeel, H., & Vanschoren, J. (2007). Experiment databases: Towards an improved experimental methodology in machine learning. In J. N. Kok, J. Koronacki, R. L. de Mantaras, S. Matwin, D. Mladenič, & A. Skowron (Eds.), *European conference on knowledge discovery in databases: PKDD* (pp. 6–17). DBLP.
Bong, M. (2008). Effects of parent–child relationships and classroom goal structures and motivation, help–seeking avoidance, and cheating. *Journal of Experimental Education*, *76*, 191–217. https://doi.org/10.3200/JEXE.76.2.191-217
Brandes, B. (1986). *Academic honesty: A special study of California students*. California State Department of Education, Bureau of Publications.
Bronfenbrenner, U. (1993). The ecology of cognitive development: Research models and fugitive findings. In R. H. Wozniak & K. W. Fischer (Eds.), *Development in context: An introduction* (pp. 3–44). Academic Press.
Bruer, K. C., Zanette, S., Ding, X. P., Lyon, T. D., & Lee, K. (2019). Identifying liars through automatic decoding of children's facial expressions. *Child Development*, *91*, e995–e1011. https://doi.org/10.1111/cdev.13336
Bucciol, A., Cicognani, S., & Montinari, N. (2017). *Cheating in academia: The relevance of social factors*. Working papers 15/2017, University of Verona, Department of Economics.
Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists. *New Directions for Program Evaluation*, *31*, 67–77. https://doi.org/10.1002/ev.1434
Chan, K., Lee, T. W., Sample, P. A., Goldbaum, M. H., Weinreb, R. N., & Sejnowski, T. J. (2002). Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Transactions on Biomedical Engineering*, *49*, 963–974.
Cizek, G. C. (1999). *Handbook of educational policy*. Academic Press.
Cochran, J. K. (2015). The effects of life domains, constraints, and motivations on academic dishonesty: A partial test and extension of Agnew's general theory. *International Journal of Offender Therapy and Comparative Criminology*, *61*, 1288–1308. https://doi.org/10.1177/0306624X15618689
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
Cuadrado, D., Salgado, J. F., & Moscoso, S. (2019). Prevalence and correlates of academic dishonesty: Towards a sustainable university. *Sustainability*, *11*, 1–20. https://doi.org/10.3390/su11216062
Davy, J. A., Kincaid, J. F., Smith, K. J., & Trawick, M. A. (2007). An examination of the role of attitudinal characteristics and motivation on the cheating behavior of business students. *Ethics & Behavior*, *17*, 281–302. https://doi.org/10.1080/10508420701519304
Desalegn, A., & Berhan, A. (2014). Cheating on examinations and its predictors among undergraduate students at Hawassa University College of Medicine and Health Science, Hawassa, Ethiopia. *BMC Medical Education*, *14*, 1–11. https://doi.org/10.1186/1472–6920–14–89
Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, *17*, 1101–1119. https://doi.org/10.1177/1745691621103767
Ding, X. P., Omrin, D. S., Evans, A. D., Fu, G., Chen, G., & Lee, K. (2014). Elementary school children's cheating behavior and its cognitive correlates. *Journal of Experimental Child Psychology*, *121*, 85–95. https://doi.org/10.1016/j.jecp.2013.12.005
Drubin, D. G. (2015). Great science inspires us to tackle the issue of data reproducibility. *Molecular Biology of the Cell*, *26*, 3679–3680. https://doi.org/10.1091/mbc.E15-09-0643
Gao, C., Sun, H., Wang, T., Tang, M., Bohnen, N. I., Müller, M. L., Herman, T., Giladi, N., Kalinin, A., Spino, C., Dauer, W., Hausdorff, J. M., & Dinov, I. D. (2018). Model–based and model–free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. *Scientific Reports*, *8*, 7129. https://doi.org/10.1038/s41598-018-24783-4
Ghanem, C. M., & Mozahem, N. A. (2019). A study of cheating beliefs, engagement, and perception—The case of business and engineering students. *Journal of Academic Ethics*, *17*, 291–312. https://doi.org/10.1007/s10805-019-9325-x
Ghorbani, A., & Zou, J. (2019). *Data shapley: Equitable valuation of data for machine learning*. Working paper. PMLR.
Golino, H. F., Gomes, C. M. A., & Andrade, D. (2014). Predicting academic achievement of high–school students using machine learning. *Psychology*, *5*, 2046–2057. https://doi.org/10.4236/psych.2014.518207
Haines, V. J., Diekhoff, G. M., LaBeff, E. E., & Clark, R. E. (1986). College cheating: Immaturity, lack of commitment, and the neutralizing attitude. *Research in Higher Education*, *25*, 342–354. https://doi.org/10.1007/BF00992130
Hartshorne, H., & May, M. A. (1928). *Studies in deceit*. McMillan.
Heyman, G. D., Fu, G., Lin, J., Qian, M. K., & Kang, L. (2015). Eliciting promises from children reduces cheating. *Journal of Experimental Child Psychology*, *139*, 242–248. https://doi.org/10.1016/j.jecp.2015.04.013
Hrabak, M., Vujaklija, A., Vodopivec, I., Hren, D., Marušić, M., & Marušić, A. (2004). Academic misconduct among medical students in a post–communist country. *Medical Education*, *38*, 276–285. https://doi.org/10.1111/j.1365–2923.2004.01766.x
Huang, C., Yang, S. C., & Chen, A. (2015). The relationships among students' achievement goals, willingness to report academic dishonesty, and engaging in academic dishonesty. *Social Behavior and Personality*, *43*, 27–38. https://doi.org/10.2224/sbp.2015.43.1.27
Inderbitzen, H. M. (1994). *Adolescent peer social competence*. Springer US.
Ip, E. J., Nguyen, K., Shah, B. M., Doroudgar, S., & Bidwal, M. K. (2016). Motivations and predictors of cheating in pharmacy school. *American Journal of Pharmaceutical Education*, *80*, 133. https://doi.org/10.5688/ajpe808133
Ives, B., Alama, M., Mosora, L. C., Mosora, M., Grosu-Radulescu, L., Clinciu, A. I., Cazan, A. M., Badescu, G., Tufis, C., Diaconu, M., & Dutu, A. (2017). Patterns and predictors of academic dishonesty in Romanian university students. *Higher Education*, *74*, 815–831. https://doi.org/10.1007/s10734-016-0079-8
Ives, B., & Giukin, L. (2020). Patterns and predictors of academic dishonesty in Moldovan university students. *Journal of Academic Ethics*, *18*, 71–88. https://doi.org/10.1007/s10805-019-09347-z
Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology*, *27*, 209–228. https://doi.org/10.1006/ceps.2001.1088
Just, M. A., Pan, L., Cherkassky, V. L., McMakin, D. L., Cha, C., Nock, M. K., & Brent, D. (2017). Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nature Human Behaviour*, *1*, 911–919. https://doi.org/10.1038/s41562-017-0234-y
Kafle, A., & Thakali, M. (2013). *Social relations in adolescence: Roles of parents and peer relationship in adolescent psychosocial development*. (Bachelor thesis, degree programme in nursing). Kemi–Tornio University of Applied Sciences.
Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., Ebert, D. D., Hwang, I., Li, J., de Jonge, P., Nierenberg, A. A., Petukhova, M. V., Rosellini, A. J., Sampson, N. A., Schoevers, R. A., Wilcox, M. A., & Zaslavsky, A. M. (2016). Testing a machine–learning algorithm to predict the persistence and severity of major depressive disorder from

baseline self–reports. *Molecular Psychiatry*, *21*, 1366–1371. https://doi.org/10.1038/mp.2015.198

Koscielniak, M., & Bojanowska, A. (2019). The role of personal values and student achievement in academic dishonesty. *Frontiers in Pediatrics*, *10*, 1887. https://doi.org/10.3389/fpsyg.2019.01887

Krou, M. R., Fong, C. J., & Hoff, M. A. (2021). Achievement motivation and academic dishonesty: A meta–analytic investigation. *Educational Psychology Review*, *33*, 427–458. https://doi.org/10.1007/s10648-020-09557-7

Kurt, I., Ture, M., & Kurum, A. T. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications*, *34*, 366–374. https://doi.org/10.1016/j.eswa.2006.09.004

Lambert, E. G., Hogan, N. L., & Barton, S. M. (2003). Collegiate academic dishonesty revisited: What have they done, how often have they done it, who does it, and why did they do it. *Electronic Journal of Sociology*, *7*, 1–27.

Lee, K. (2013). Little liars: Development of verbal deception in children. *Child Development Perspectives*, *7*, 91–96. https://doi.org/10.1111/cdep.12023

Lee, K., & Evans, A. D. (2013). Becoming a moral relativist: Children's moral conceptions of honesty and dishonesty in different social contexts. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 361–366). Oxford University Press.

Lee, S. D., Kuncel, N. R., & Gau, J. (2020). Personality, attitude, and demographic correlates of academic dishonesty: A meta–analysis. *Psychological Bulletin*, *146*, 1042–1058. https://doi.org/10.1037/bul0000300

Lim, V. K. G., & See, S. K. B. (2001). Attitudes toward, and intentions to report, academic cheating among students in Singapore. *Ethics & Behavior*, *11*, 261–274. https://doi.org/10.1207/S15327019EB1103_5

Livieris, I. E., Kotsilieris, T., Dimopoulos, I., & Pintelas, P. (2018). Decision support software for forecasting patient's length of stay. *Algorithms*, *11*, 199. https://doi.org/10.3390/a11120199

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

Marsden, H., Carroll, M., & Neill, J. T. (2005). Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. *Australian Journal of Psychology*, *57*, 1–10. https://doi.org/10.1080/00049530412331283426

McCabe, D. L., & Abdallah, F. H. (2008). Academic dishonesty in the Middle East: Individual and contextual factors. *Research in Higher Education*, *49*, 451–467.

McCabe, D. L., Butterfield, K. D., & Trevino, L. K. (2012). *Cheating in college: Why students do it and what educators can do about it*. John Hopkins University Press.

McCabe, D. L., & Treviño, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education*, *64*, 522–538.

McCabe, D. L., & Treviño, L. K. (1997). Individual and contextual influences on academic dishonesty: A multicampus investigation. *Research in Higher Education*, *38*, 379–396. https://doi.org/10.1023/A:1024954224675

Megehee, C. M., & Spake, D. F. (2008). The impact of perceived peer behavior, probable detection and punishment severity on student cheating behavior. *Marketing Education Review*, *18*, 5–19. https://doi.org/10.1080/10528008.2008.11489033

Meiseberg, B., Ehrmann, T., & Prin, A. (2016). "Anything worth winning is worth cheating for"? Determinants of cheating behavior among business and theology students. *Journal of Business Economics*, *87*, 985–1016. https://doi.org/10.1007/s11573-016-0842-4

Michaels, J. W., & Miethe, T. D. (1989). Applying theories of deviance to academic cheating. *Social Science Quarterly*, *70*, 870.

Miller, A., Shoptaugh, C., & Wooldridge, J. (2011). Reasons not to cheat, academic–integrity responsibility, and frequency of cheating. *Journal of Experimental Education*, *79*, 169–184. https://doi.org/10.1080/00220970903567830

Murdock, T. B., & Anderman, E. M. (2006). Motivational perspectives on student cheating: Toward an integrated model of academic dishonesty. *Educational Psychologist*, *41*, 129–145. https://doi.org/10.1207/s15326985ep4103_1

Nathanson, C., Paulhus, D. L., & Williams, K. M. (2006). Predictors of a behavioral measure of scholastic cheating: Personality and competence but not demographics. *Contemporary Educational Psychology*, *31*, 97–122. https://doi.org/10.1016/j.cedpsych.2005.03.001

Newstead, S. E., Franklyn-Stokes, A., & Armstead, P. (1996). Individual differences in student cheating. *Journal of Educational Psychology*, *88*, 229–241. https://doi.org/10.1037/0022-0663.88.2.229

Ng, A. (2017). *Machine learning yearning*. On-Line Press. https://github.com/ajaymache/machine-learning-yearning

O'Connor, A. M., & Evans, A. D. (2019). The role of theory of mind and social skills in predicting children's cheating. *Journal of Experimental Child Psychology*, *179*, 337–347. https://doi.org/10.1016/j.jecp.2018.11.018

Oran, N. T., Can, H. O., Senol, S., & Hadimli, A. P. (2016). Academic dishonesty among health science school students. *Nursing Ethics*, *23*, 919–931. https://doi.org/10.1177/0969733015583929

O'Rourke, J., Barnes, J., Deaton, A., Fulks, K., Ryan, K., & Rettinger, D. A. (2010). Imitation is the sincerest form of cheating: The influence of direct knowledge and attitudes on academic dishonesty. *Ethics & Behavior*, *20*, 47–64. https://doi.org/10.1080/10508420903482616

Özcan, M., Yeniçeri, N., & Çekiç, E. G. (2019). The impact of gender and academic achievement on the violation of academic integrity for medical faculty students, a descriptive cross–sectional survey study. *BMC Medical Education*, *19*, 427. https://doi.org/10.1186/s12909-019-1865-7

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Vanderplas, J. (2011). Scikit–learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Popoola, I. T., Garner, B., Ammeter, A., Krey, N., Beu Ammeter, D., & Schafer, S. (2017). How does ethics institutionalization reduce academic cheating? *Journal of Education for Business*, *92*, 29–35. https://doi.org/10.1080/08832323.2016.1274710

Pulvers, K., & Diekhoff, G. M. (1999). The relationship between academic dishonesty and college classroom environment. *Research in Higher Education*, *40*, 487–498.

Putarek, V., & Pavlin-Bernardic, N. (2020). The role of self–efficacy for self–regulated learning, achievement goals, and engagement in academic cheating. *European Journal of Psychology of Education*, *35*, 647–671. https://doi.org/10.1007/s10212-019-00443-7

Ramberg, J., & Modin, B. (2019). School effectiveness and student cheating: Do students' grades and moral standards matter for this relationship? *Social Psychology of Education*, *22*, 517–538. https://doi.org/10.1007/s11218-019-09486-6

Rettinger, D. A., & Kramer, Y. (2009). Situational and personal causes of student cheating. *Research of Higher Education*, *50*, 293–313. https://doi.org/10.1007/s11162-008-9116-5

Shapley, L. S. (1953). A value for n–person games. *Contributions to the Theory of Games*, *2*, 307–317.

Simpson, E., & Yu, K. (2012). Closer to the truth: Electronic records of academic dishonesty in an actual classroom setting. *Ethics & Behavior*, *22*, 400–408. https://doi.org/10.1080/10508422.2012.702514

Siniver, E. (2013). Cheating on exams: The case of Israeli students. *College Student Journal*, *47*, 593–604.
Smith, M., & Alvarez, F. (2021). Identifying mortality factors from machine learning using Shapley values—A case of Covid19. *Expert Systems with Applications*, *178*, 114832. https://doi.org/10.1016/j.eswa.2021.114832
Stephens, J. M. (2017). Bridging the divide: The role of motivation and self–regulation in explaining the judgment–action gap related to academic dishonesty. *Frontiers in Psychology*, *9*, 246. https://doi.org/10.3389/fpsyg.2018.00246
Stiles, B. L., Wong, N. C. W., & LaBeff, E. E. (2018). College cheating thirty years later: The role of academic entitlement. *Deviant Behavior*, *39*, 823–834. https://doi.org/10.1080/01639625.2017.1335520
Stylianou, N., Akbarov, A., Kontopantelis, E., Buchan, I., & Dunn, K. W. (2015). Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches. *Burns*, *41*, 925–934. https://doi.org/10.1016/j.burns.2015.03.016
Sufriyana, H., Husnayain, A., Chen, Y. L., Kuo, C. Y., Singh, O., Yeh, T. Y., Wu, Y. W., & Su, E. C. Y. (2020). Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: Systematic review and meta-analysis. *JMIR Medical Informatics*, *8*, e16503. https://doi.org/10.2196/16503
Sun, Y. H., Liu, Q. J., Lee, N. Y., Li, X. H., & Lee, K. (2022). A novel machine learning approach to shorten depression risk assessment for convenient uses. *Journal of Affective Disorders*, *312*, 275–291. https://doi.org/10.1016/j.jad.2022.06.035
Sun, Y. H., Luo, H., & Lee, K. (2022). A novel approach for developing efficient and convenient short assessments to approximate a long assessment. *Behavior Research Methods*, *54*, 2802–2828. https://doi.org/10.3758/s13428-021-01771-7
Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.
Turnipseed, L. D., & Landay, K. (2018). The role of the dark triad in perceptions of academic incivility. *Personality and Individual Differences*, *135*, 286–291. https://doi.org/10.1016/j.paid.2018.07.029
Voelker, P. F. (1921). *The function of ideals and attitudes in social education: An experimental study (no. 112)*. Teachers College, Columbia University.
Volpe, R., Davidson, L., & Bell, M. C. (2008). Faculty attitudes and behaviors concerning student cheating. *College Student Journal*, *42*, 164–175.
Waltzer, T., & Dahl, A. (2020). Students' perceptions and evaluations of plagiarism: Effects of text and context. *Journal of Moral Education*, *50*, 1–16. https://doi.org/10.1080/03057240.2020.1787961
Waltzer, T., & Dahl, A. (2022). The moral puzzle of academic cheating: Perceptions, evaluations, and decisions. In D. Rettinger & T. Bertram Gallant (Eds.), *Cheating and academic integrity: Lessons from 30 years of research*. Wiley/Jossey Bass.
Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education*, *39*, 235–274. https://doi.org/10.2307/40196379
Whitley, B. E., Nelson, A. B., & Jones, C. J. (1999). Gender differences in cheating attitudes and classroom cheating behavior: A meta–analysis. *Sex Roles*, *41*, 657–680. https://doi.org/10.1023/A:1018863909149
Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*, 1100–1122. https://doi.org/10.1177/1745691617693393
Youmans, R. J. (2011). Does the adoption of plagiarism–detection software in higher education reduce plagiarism? *Studies in Higher Education*, *36*, 749–761. https://doi.org/10.1080/03075079.2010.523457
Yu, H. W., Glanzer, P. L., Sriram, R., Johnson, B. R., & Moore, B. (2016). What contributes to college students' cheating? A study of individual factors. *Ethics & Behavior*, *27*, 401–422. https://doi.org/10.1080/10508422.2016.1169535
Zanette, S., Gao, X. Q., Brunet, M., Bartlett, M. S., & Lee, K. (2016). Automated decoding of facial expressions reveals marked differences in children when telling antisocial versus prosocial lies. *Journal of Experimental Child Psychology*, *150*, 165–179. https://doi.org/10.1016/j.jecp.2016.05.007
Zhao, L., Heyman, G. D., Chen, L. L., & Lee, K. (2018). Telling young children they have a reputation for being smart promotes cheating. *Developmental Science*, *21*, e12585. https://doi.org/10.1111/desc.12585
Zhao, L., Mao, H. Y., Compton, B. J., Peng, J. J., Fu, G., Fang, F., Heyman, G. D., & Lee, K. (2022). Academic dishonesty and its relations to peer cheating and culture: A meta-analysis of the perceived peer cheating effect. *Educational Research Review*, *36*, 100455. https://doi.org/10.1016/j.edurev.2022.100455
Zhao, L., Zheng, Y., Compton, B. J., Qin, W., Zheng, J., Fu, G., Lee, K., & Heyman, G. D. (2020). The moral barrier effect: Real and imagined barriers can reduce cheating. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 19101–19107. https://doi.org/10.1073/pnas.2002249117
Zhao, L., Zheng, Y., Mao, H. Y., Compton, B. J., Fu, G., Heyman, G. D., & Lee, K. (2021). Using environmental nudges to reduce academic cheating in young children. *Developmental Science*, *24*, e13108. https://doi.org/10.1111/desc.13108

How to cite this article: Zhao, L., Zheng, Y., Zhao, J., Li, G., Compton, B. J., Zhang, R., Fang, F., Heyman, G. D., & Lee, K. (2023). Cheating among elementary school children: A machine learning approach. *Child Development*, *94*, 922–940. https://doi.org/10.1111/cdev.13910

APPENDIX

[The following is the complete set of survey items. Notes appear in brackets.]

Directions: The survey is anonymous. There are no right or wrong answers, so just answer as honestly as possible. Thank you for your participation.

Q1. Have you ever engaged in cheating during an exam?

[This item used a five–point scale: 1 = never, 2 = somewhat infrequently, 3 = neither frequently nor infrequently, 4 = somewhat frequently, 5 = extremely frequently].

Q2. To what degree do you think cheating on exams is common among your classmates?

[This item used a five–point scale: 1 = not common at all, 2 = somewhat uncommon, 3 = neither common nor uncommon, 4 = somewhat common, 5 = extremely common].

Q3. How often do you think your classmates have engaged in cheating during an exam?

[This item used a five–point scale: 1 = never, 2 = somewhat infrequently, 3 = neither frequently nor infrequently, 4 = somewhat frequently, 5 = extremely frequently].

Q4. How often do you think your classmates have engaged in each the following forms of behavior during an exam?

[The following items used a five–point scale: 1 = never, 2 = somewhat infrequently, 3 = neither frequently nor infrequently, 4 = somewhat frequently, 5 = extremely frequently].

1. Bringing unauthorized materials to an exam.
2. Copying answers from a textbook during an exam.
3. Passing notes during an exam.
4. Copying answers from a neighbor during an exam.
5. Using tools (e.g., dictionaries, cellphones, or smart watches) to search for answers during an exam.
6. Working with one or more classmates to share answers during an exam
7. Deliberately giving a classmate a wrong answer during an exam.
8. Secretly changing a test score.

Q5. To what degree is cheating on exams acceptable to you?

[This item used a five–point scale: 1 = not acceptable at all, 2 = somewhat unacceptable, 3 = neither acceptable nor unacceptable, 4 = somewhat acceptable, 5 = completely acceptable] nothing special.

Q6. To what degree do you think cheating on exams is acceptable among your classmates?

[This item used a five–point scale: 1 = not acceptable at all, 2 = somewhat unacceptable, 3 = neither acceptable nor unacceptable, 4 = somewhat acceptable, 5 = completely acceptable].

Q7. To what degree do you think each of the following items can serve as an effective way to reduce cheating on exams?

[The following items used a five–point scale: 1 = not effective at all, 2 = somewhat ineffective, 3 = neither effective nor ineffective, 4 = somewhat effective, 5 = extremely effective].

1. Increasing the severity of consequences of getting caught cheating (e.g., giving a zero score).
2. Students who sit next to each other getting different versions of the test.
3. Working harder.
4. Teachers emphasizing that academic cheating represents a serious moral transgression
5. Having a teacher who is greatly liked by their students to teach the class.
6. Teachers giving sharp criticism or punishment.
7. Teachers recognize classroom role models by giving praise or prizes to students who behave honestly on exams
8. Parents being informed when their children are caught cheating, and the parents in turn giving sharp criticism or punishment

Q8. To what degree do you think each of the following items is severe?

[The following items used a five–point scale: 1 = not severe at all, 2 = less severe, 3 = neither severe nor non-severe, 4 = somewhat severe, 5 = extremely severe].

1. Being criticized by one's teacher.
2. Being punished by one's teacher.
3. Being criticized by one's parents.
4. Being punished by one's parents.
5. Being criticized or rejected by one's classmates.

Q9. School: _____.

Q10. Date of Birth:_____.

Q11. Gender: _____.

Q12. Grade: _____.

Q13. Do you have any siblings?

[Respondents selected a single item from the above set of options.]

1. No, I am the only child.
2. Yes, I have one or more older siblings.
3. Yes, I have one or more younger siblings.
4. Yes, I have both older and younger siblings

Q14. What level do you think your academic performance is in the class?

[Respondents selected a single item from the above set of options.]

1. Above the average.
2. Average.
3. Below the average.