MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2025, 21(6), em2646
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/16506

Understanding teacher decisions about student grade level promotions

Maria Kosogorova [1*], Sona Antonyan [1], Manasi Singhal [1], Jenya Brodskaia [1], Anna Mirny [1], Oksana Ilina [1], Ishan Jetwani [1], Yamato Murakami [1], Mikhail Pevunov [1], Sourav Sinha [1], Vidhyuth Venugopal [1], Larry Ludlow [2]

[1] Russian School of Mathematics, Newton, MA, USA
[2] Boston College, Boston, MA, USA

Received 24 November 2024 ▪ Accepted 19 May 2025

Abstract
Existing research indicates a potential correlation between students' online homework and classwork behaviors and their educational progress. Exploring this link could help identify key performance behaviors that are essential for improvement. Ideally, encouraging students to enhance these behaviors would lead to better learning outcomes, faster academic progress, and greater overall educational potential. In this study, the online homework and classwork behaviors of fourth-grade students enrolled in an after-school mathematics program were analyzed to predict their academic placement for the following year, which could remain the same, move up a level, or drop down. The results from several predictive models confirm that there is indeed a correlation between students' online behaviors and their future academic placement. Based on these findings, recommendations are proposed to help students boost their progress potential. Moreover, distinct suggestions are offered for both new and returning students.

Keywords: process data, predicting promotions, mathematical achievement, after-school programs

INTRODUCTION

Educational research in the area of student academic performance and potential relies heavily on quantitative procedures. Establishing statistical tendencies helps researchers understand the reasons underlying different academic outcomes (Zimmerman & Kitsantas, 2005) in order to develop effective teaching strategies (Rosário et al., 2015; Roschelle et al., 2016), preferably tailored to different student groups.

Previous research has revealed homework behavior characteristics that are crucial for student success (Fan et al., 2017; Valle et al., 2016), but this line of research has typically focused on improving teacher behaviors (e.g., having teachers generate more detailed observation reports about student problem behaviors) (Algozzine et al., 2010; Fateen & Mine, 2021). The other half of the educational process, however, is student involvement and effort, and this is what we, as teaching practitioners, seek to improve.

It is also argued that homework policies and practices at the elementary school level can influence the development of student skills and behaviors necessary for success at the secondary level (Epstein, 1983), aid the retention of factual information, enhance information processing, and develop greater self-discipline (Ramdass & Zimmerman, 2011; Wu et al., 2023). All these factors contribute to the learning potential of a student (Oppong et al., 2018). Hence, our research program has focused on determining what practices and habits students need to enhance in order to achieve their fullest potential (Ramdass & Zimmerman, 2011; Wu et al., 2023).

The outcome for assessing "learning potential" is often simply categorized as the "course pass/course fail" dichotomy (Fernandes et al., 2018). There is,

✉ maria.kosogorova@mathschool.com **(*Correspondence)** ✉ sona.antonyan@russianschool.com
✉ manasi.singhal@russianschool.com ✉ jenya.brodskaia@russianschool.com ✉ anna.mirny@mathschool.com
✉ oksana.ilina@russianschool.com ✉ ishjet32@gmail.com ✉ yamato.murakami@gmail.com ✉ mikhail.pevunov@gmail.com
✉ sinhasourav2006@gmail.com ✉ vidhu16.v@gmail.com ✉ ludlow@bc.edu

Contribution to the literature

- This study contributes to the literature by addressing a notable gap in predictive modeling research at the elementary school level, where process-based data remains largely underutilized.
- By applying and comparing three complementary models–logistic regression (LR), linear discriminant analysis (DA), and decision trees (DT)–on multi-year Russian School of Mathematics (RSM) process data, we demonstrate the value of behavioral indicators in forecasting categorical learning outcomes.
- We provide insights for early intervention with the focus on developing productive work habits.

however, research with the outcome defined as promotion to the next level (Delen, 2010), which, we believe, reflects student learning potential more appropriately. Passing a course demonstrates success in that particular course, whereas promotion is a sign that the student can continue progress at a higher level (King et al., 2008).

Even though most of this line of educational research has been conducted using demographic data, such as student gender, income level, parents' level of education etc. (Raj & Vannan 2020), recent studies suggest that this approach is less useful than using actual educational data (Chavez et al., 2023). Some studies, for example, have used grades and other academic achievements (quizzes, tests, and GPA) as predictors for academic success (Botchway et al., 2024; Zabriskie et al., 2019) but grades are a direct criterion for passing or failing the course (Alshanqiti & Namoun, 2020). In contrast, there is a growing body of educational research using student online process data, which includes recording student behaviors in homework and classwork recorded online with timestamps (Feng & Roschelle, 2016; Mogessie et al., 2015; Namoun & Alshanqiti, 2020). When process data analysis procedures are employed in an educational setting the process has been referred to as educational data mining (Alhassan et al., 2020).

Most of the current educational data mining has been conducted on college or high school students (Cruz Jesus et al., 2020). There has been research, however, with seventh-grade process data to predict standardized test scores (Feng & Roschelle, 2016) and fifth-grade educational data mining to track online homework activities and their correlation with achievement (Shechtman et al., 2019). Another recent study accumulated fourth grade homework logs in connection with learning achievement (Wu et al., 2023).

These data mining studies suggest that there might exist a correlation between a student's homework/classwork online behaviors and their potential for educational progress. Further evidence of such a correlation may result in highlighting the performance behaviors crucial for progress. Ideally, enhancements of these behaviors on behalf of a student would lead to improvements in learning pace, academic results and educational potential.

Predicting categorical learning outcomes (such as pass/fail) based on student process data has proven highly effective in secondary and post-secondary education (Arizmendi et al., 2023; Doctor, 2023). These models are often used to improve student performance by identifying and addressing weaknesses in work habits, such as procrastination, inconsistent engagement, or low participation. Moreover, the integration of multiple predictive algorithms has been shown to reduce misclassification rates and increase the reliability of interventions (Karalar & Gürüler, 2021).

Surprisingly, however, this approach remains underutilized in research concerning elementary school students. Existing studies at this level predominantly rely on static predictors such as demographic and physical information (Wickramasinghe et al., 2024) or standardized test scores (Cornell-Farrow & Garrard, 2020). This represents a significant oversight. Early schooling is a formative period when students' learning behaviors and self-regulation skills begin to take shape–habits that are foundational for future academic trajectories. Ignoring process-based predictors at this critical stage means missing a crucial opportunity for early intervention. While demographic predictors can identify broad risk factors, they do not capture the day-to-day behavioral patterns that more directly influence success or failure (Kassarnig et al., 2018). In effect, research has not only overlooked a powerful set of predictive indicators but has also limited its scope at a moment when timely behavioral insights could be most impactful and easily acted upon.

Our study addresses this gap by leveraging process data to predict categorical outcomes in elementary students, focusing specifically on indicators of work habits. We extend the methodological range by employing three complementary models–LR, linear DA, and DT–and analyzing cumulative data collected over a three-year period to improve the robustness of prediction. Ultimately, our goal is not only to model outcomes but to translate these insights into actionable recommendations that help students cultivate effective work habits in elementary school–setting a foundation for academic success in middle school and beyond.

Context

This study was conducted by a group of RSM high school students guided by their RSM teachers and a university faculty advisor. This work is part of a multi-year research-based educational project created and conducted by RSM. RSM is an after-school program focused on delivering mathematics education in grades K-12 (Russian School of Mathematics, 2020). The RSM philosophy of teaching is based on the Vygotskian approach to child development and learning, emphasizing the collaborative nature of learning by the construction of knowledge through social negotiation (Vygotsky, 1926). This approach seeks to give students skills through student-peer and student-teacher interactions while fostering their own problem-solving abilities.

Students who enroll in RSM are initially sorted into three levels of program (from least to most difficult): accelerated, advanced, and honors. This initial placement is made by a principal or an experienced teacher after an initial evaluation of a child's skills and potential. The accelerated level is designed for new students, focusing on teaching basic RSM practices, filling gaps, and adapting to the new curriculum. The advanced level is the 'golden standard' of RSM, where most students gain solid mathematical knowledge and where very little time is dedicated to basic skills and working habits development. The honors level is a very special level of teaching which is not suitable for everyone due to its extremely fast pace and high expectations of students' independence.

At the end of each year, teachers assess the students' potential and performance and suggest placements for the next academic year. The assessment includes classwork and homework performance, test scores, overall capability for independent and group work, and attitude. This end-of-year placement generally presumes that every fourth-grade student is promoted to grade 5. In most cases, accelerated-level students are promoted to advanced if they have mastered the basics, or stay in accelerated if they have not. Students in advanced are unlikely to be promoted to honors, unless they show exceptional mathematical proficiency and potential, but are also unlikely to be demoted to accelerated, unless there is a very significant structural gap discovered by a teacher. Honors students have a chance to be demoted to the advanced level if a teacher decides that they need a slower pace or less challenge, otherwise they stay in honors.

RSM is a mainly in-person educational program, but it features a unique online platform, which, among other functions, allows students from grade 3 and up to submit answers for homework online and receive instant feedback. The homework portal allows the teacher to see each attempt of each student and overall class statistics before the beginning of each class and adapt homework review accordingly. This approach combines online and in-person education for a smoother delivery of material in a blended mode. The platform also keeps track of students' progress by summarizing attendance, grades and scores, report card data, teachers' notes and written homework submitted through the portal.

This study was conducted during the third year of our ongoing project-based research program. The first years' study revealed how homework scores and time spent on homework were affected by the COVID-19 environment (Ilina et al., 2023). During the second year of the project the next cohort of students expanded the study of homework practices by using a set of online "process data" behaviors to predict homework scores and summative test scores (Ilina et al., In press). The current study, conducted by the third cohort of students, delved deeper into understanding how grade 4 elementary students' online behaviors impact their learning and subsequent educational progress, i.e., promotion to one of the grade 5 curriculum levels.

Similar to the first two cohorts, the student collaborators participated in the literature review, data analysis and interpretation of results. However, as an extension of our evolving research-based curriculum, the students also suggested new directions of analysis, helped formulate the research conclusions and their significance, and finished their coursework with a draft of a conference proposal for 2025 and a draft of a publishable manuscript.

The purpose of this study was to determine which of our extensive list of process data variables may be underlying teacher recommendations and decisions for maintaining or changing the curriculum level of a student in the next academic year. A change in curriculum level reflects a change in a student's learning momentum, or progress potential, which, as we claim, should be possible to predict by a combination of a student's learning habits.

Research Questions

RQ1. How well do a combination of teacher ratings and student-related process variables predict next year (grade 5) curriculum level placement for RSM grade 4 students?

RQ2. Does the correlation of the set of process variables to the predicted grade 5 level remain consistent for two different levels of grade 4 students?

RQ3. Are there any consistent placement misprediction tendencies, and what could be the reasons behind them?

METHOD

Sample

The sample consists of approximately 2000 grade 4 RSM students enrolled in our Online branch (no in-person classes) in the 2020-2021, 2021-2022 and 2022-2023 academic years. Student data from the 2023-2024 academic year was used for cross-validation predictive purposes. We chose grade 4 over other grades for the following reasons:

(a) in 2020-2023 this was the only elementary level grade that used the online homework system,

(b) we prefer studying at the elementary level because, unlike other levels, their homework is uniform throughout all classes, and a teacher cannot change it for one particular class, whereas for older grades the teachers can change the contents of homework,

(c) fourth grade curricula did not change in 2020-2023, and

(d) there are few, if any, studies of elementary school mathematics achievement taken across several years using online studying and learning behaviors.

The students in our sample have a mean age of approximately 10.2 years, ranging from 9 to 12 years old. The students are mostly from the United States with English being the teaching language.

Research Design

We employed a retrospective quantitative, non-manipulative, multivariate multi-method design (Cooley & Lohnes, 1971). In this design we approached the matrix of predictor-by-student archival data as a p-dimensional multivariate geometric space (Wickens, 1995). The p dimensions correspond to the number of predictors collected on each student. Each student then has a fixed coordinate point in this p space. One's understanding of the associations among this swarm of points (Cooley & Lohnes, 1971, p. 263) depends on the procedures used to investigate this space. For example, many multivariate applications take a single procedure, such as factor analysis or multivariate analysis of variance, to investigate relations among the variables or how the variables differentiate among various groups (Cooley & Lohnes, 1971).

Other studies, however, investigate the extent to which various statistical procedures yield different, yet complementary, insights into variable and group relationships (Ludlow, 1999; Ludlow et al., 2017). Our associative data analysis strategy (Green, 1976), therefore, was to apply three different multivariate procedures to see the extent to which their analyses agreed yet revealed unique insights into the promotion decisions that teachers made. Furthermore, we employed a "simultaneous" forced-entry estimation approach to enable comparisons across the three models based on the same predictors (Hair et al., 1987).

Data

Each RSM course lasts 36 weeks with one class per week. Students are expected to complete an hour-long homework assignment after each class, making it 35 assignments. The number of problems in each assignment is different (but averages 14) and consists of problems similar to the ones solved and explained in class. The answers to homework problems need to be submitted online (graded right/wrong with partial credit for second-attempt binary questions and too many attempts). Additionally, the solutions for homework are submitted on paper (by uploading a photo), for a separate credit.

Proprietary RSM software collects student interactions with the homework portal and stores it in the database. We do not collect clicks and change of tabs, but we can retrieve the timestamps, values and correctness of every student's attempt, their scores, hint usage, as well as teacher input such as grades and report card values. In order to avoid overfitting the model with all the available RSM software-generated indicator predictors possible (resulting in unrealistically large coefficients and standard errors), an initial consideration of the educationally substantive merit of each predictor was undertaken. Our review, guided by the literature, focused on practical behaviors that can be encouraged in students by teachers and parents. Some of the predictors we selected are frequently used for this type of student performance research (e.g., average time spent on homework, average attempts per homework problem, percentage of correct answers, percentage of assignments completed on time (Bezirhan et al., 2020; Feng & Roschelle, 2016)). This step was followed by univariate analyses to establish the initial presence of a relationship between the potential predictors and the outcomes (Hosmer & Lemeshow, 1989).

In addition, we retrieved two types of classwork performance evaluation by the teacher from the database, they are participation and behavior predictors. A teacher evaluates student behavior as a combination of respectfulness and adhering to classroom rules. Participation is evaluated as student ability and willingness to participate in discussions, volunteer answers and ideas, and contribute during teamwork time. These evaluations are recorded as integers ranging from 1 to 5, with 1 being the least possible value and 5 corresponding to the best participation/behavior. The scores are averaged across two reporting periods and are then categorized from 1 to 5 with a step of 0.5. The final set of student online process data and teacher-generated behavior indicators used to model student placement in the next academic year is presented in **Table 1**.

Data Cleaning and Outliers

Multivariate data sets need to be checked for outliers. At the univariate level, "discordant" and "contaminant" observations (Beckman & Cook, 1983) on a predictor may arise from many sources including, among others, data file entry errors, unreliable measures, and legitimate instances of unexpected responses (Kruskal, 1960). The immediate impact of such unusual data points is to skew measures of central tendency and inflate

Table 1. List of predictors used

Predictors	Description	Data transformations*
avg_attempts_per_hw_problem	Total number of times an answer was submitted to any student's HW problem divided by the total number of problems in all HW assignments *Indicator of: persistence*	Square root
average_days_before_first_attempt	The number of days before a student submits a first attempt on homework since the date it's assigned. The date of first submission is taken and then the date of assignment is subtracted from it, and a number between 1 to 7 is calculated. *Indicator of: good homework habits*	Square root
fraction_hws_uploaded_out_of_opened_hws	There are 35 homework assigned, and for each one, the students can upload papers and homework to the portal. Then the number of homework uploaded is divided by 35, then converted to a percentage. *Indicator of: good homework habits*	No transformation applied
average_time_spent_on_hw_assignment	Average time spent on solving a homework assignment. When time is calculated, periods of inactivity (no answer submitted) of 15+ minutes are excluded. *Indicator of: persistence*	Square root
average_number_revisits_per_unsolved_hw_problem	If a problem wasn't solved during the first attempt - which can be several answer submissions in a row - this is how many times a student would return to it on average. A revisit is when a student worked (submitted an answer) on another problem before returning to that one. *Indicator of: consistency*	Square root
avg_number_HW_revisits	How many times a student comes back to their homework assignment on average. A revisit is coming back after a 12+ hour break. *Indicator of: consistency*	Square root
fraction_lessons_attended_excluding_first	Percent of lessons attended excluding the first lesson. Some students join later or request a different weekday at the beginning of the year due to schedule conflicts so we excluded the first lesson. *Indicator of: regular attendance*	Arcsine
participation	Average classroom participation score, based on the two semester report cards*. Scored as an integer from 1 as lowest to 5 as highest. *Indicator of: classroom work*	Square root
behavior	Average classroom behavior score, based on a set of the two semester report cards**. Scored as an integer from 1 as lowest to 5 as highest. *Indicator of: uninterrupted learning*	Square root

Note. *Where applicable & **In special circumstances, for example when a student switches classes midway through the year, there can be more than 2 scores given out per academic year

Table 2. Data statistics

Grade 4_1: Accelerated	Grade 4_2: Advanced	Grade 4_3: Honors	Total grade 4
Total students: 764 *Promoted* to grade 5_2: Advanced: 587 (77%) *Retained* at grade 5_1: Accelerated: 177 (23%)	Total students: 727 *Promoted* to grade 5_3: Honors: 91 (12.5%) *Retained* at grade 5_2: Advanced: 617 (85%) *Demoted* to grade 5_1: Accelerated: 19 (2.5%)	Total students: 336 *Retained* at grade 5_3: Honors: 256 (76%) *Demoted* to grade 5_2: Advanced: 80 (24%)	Total students: 1,827

measures of dispersion (Gnanadesikan, 1977). At the multivariate modeling level, outliers are unexpected, predicted values (Gnanadesikan, 1977). These outliers may be a consequence of univariate outliers that have influenced the estimation of the coefficients and that should, therefore, be considered for exclusion (Belsley et al., 1980; Fox, 1991). It has been suggested that this consideration should be done at either the recording stage or preliminary stage of data processing (Dixon, 1953).

Following their advice, we extracted the student data, reviewed their response records on each of the predictors, and evaluated as outliers and excluded from the analysis those records with anomalies that included:

(a) students who attended less than 10 classes and

(b) students whose online behavior seemed unrealistic and could not be reasonably explained (unreasonably long/short time periods recorded, for example).

Overall, the final pool of students has the characteristics summarized in **Table 2**.

The models we describe below worked well with the accelerated and honors level students because there were sufficient numbers of students in the two possible outcomes for these levels. The models lacked sufficient statistical power when looking at the advanced level, however, because the number of students demoted back to accelerated or promoted to honors for grade 5 is very small–even when the data were aggregated across three years. Hence, our promotion prediction analyses address only the accelerated and honors students. Because of this data limitation we consider these analyses to be a form of "proof-of-concept" to establish whether or not promotion decisions can be statistically modeled and subsequently strengthened through the collection of additional years of data and across other grade levels in the program. Additionally, for some of the analyses data transformations were performed on the predictors and outcomes.

Statistical Models

Many different forms of quantitative models have been used when predicting student outcomes, including artificial neural networks, DT, extremely randomized trees, random forest, support vector machines, K-nearest neighbors, and various forms of regression procedures (Chavez et al., 2023; Cruz Jesus et al., 2020). In contrast to strictly exploratory predictive modeling, we focused on procedures that would help us best understand, explain and reinforce the learning behaviors associated with promotion outcomes. Hence, we used DA, LR, and DT models because they provide different perspectives on the relationships among the predictors and their influence on the promotion decisions.

These models were employed in a specific order because of the additional insight they provided with each successive analysis. For example, with DA we wanted to know, among other features of the analysis, the direction of the relationship between each predictor and the outcome; LR then augmented that directional information by providing the magnitude of the odd ratios associated with each predictor; and the DT analysis revealed the numerical value along each predictor's continuum where the optimal split occurred that separated the two groups being considered (either the grade 4 accelerated students being promoted to grade 5 advanced or the grade 4 honors students being retained in the grade 5 honors level).

Since these three models are all "multivariate" in the sense that they employ multiple predictor variables in the analyses, correlated predictors may cause various statistical problems. More formally known as collinearity, multicollinearity, or ill conditioning (Belsley et al., 1980), the introduction of a second predictor into a model along with the original predictor of interest will change the estimated coefficient, standard error, significance test value, and p-value associated with the first predictor (when the two predictors are correlated). In this situation the estimated coefficient for the first predictor may be diminished or enhanced and even reversed in sign (Ludlow & Klein, 2014). As additional correlated predictors are added to a model, all previous estimates continue to change and the standard errors typically increase (Pedhazur, 1982).

Unfortunately, a simple inspection of the zero-order predictor correlation matrix cannot reveal the extent to which collinear relationships exist because the absence of high correlations is not evidence of no collinearity (Belsley et al., 1980). Collinearity is, however, present when a high multiple correlation occurs from regressing one of the predictors on the others. One way to detect collinearity is to perform an ordinary least squares (OLS) regression of the dichotomous outcome on the predictors and inspect the predictor tolerance statistics (Tol_i = 1-R-square$_i$) and the variance inflation factors ($VIF_i = 1/Tol_i$) where i is the target predictor regressed on the remaining predictors (Menard, 1995).

Since the functional form of the model for the outcome variable is irrelevant for investigating the relationship among the predictors (Menard, 1995), this procedure is appropriate for both the DA and LR models. Deleting such variables, however, runs the risk of "omitted variable bias" and one reasonable approach is to focus on the combined effects of the predictors and accept the instability of individual predictors (Menard, 1995). Since we force all predictors into the DA and LR models (a "simultaneous approach" for consistency across the results), we are less concerned with the potential effect of collinearity on the standard errors than we might be in a different study (Hair et al., 1987).

Discriminant analysis

DA is a linear model similar in its use and application to OLS regression and LR (Huberty, 1994). It differs from OLS in that the outcome variable is typically a dichotomous group (e.g., promote/retain, win/lose, recover/not recover) and the estimation criterion is to maximize differentiation between the two groups' grand means (centroids). A probability of belonging to both groups is computed (based on how close the score is to each centroid) and the student is classified into that group with a probability ≥ 0.50. Since our interest lies in the extent to which we can predict grade 5 placements using our specific set teacher and student learning predictors, this approach has been referred to as "predictive discriminant analysis" (Huberty et al., 1987, p. 307).

The standard assumptions for a two-group DA are multivariate normality and equality of the group covariance matrices (Lachenbruch, 1975). Multivariate normality, however, refers to the residuals from the outcome and its predicted values (Bock, 1975). Kendall and Stuart (1976, p. 336) assert that discrimination with two groups does not, itself, require an assumption of "multinormality". It is, however, an assumption for significance testing but Lachenbruch (1975) shows that violations typically result in a modest reduction of efficiency (larger standard errors) and accuracy of predictions.

Although Tatsuoka (1971) shows how a two-group DA reduces to a multiple regression, the typical DA has the additional assumption of equal variance-covariance matrices (Klecka, 1980). Box's (1949) M tests this assumption but is frequently ignored because its power is a function of sample size and rejection of the assumption does not necessarily imply a serious problem (Cooley & Lohnes, 1971). Still, knowing which group has greater predictor dispersion than the other may be useful as descriptive information. In either case, many researchers argue that DA is robust to violations of these assumptions (Lachenbruch, 1975).

Logistic regression model

OLS procedures applied to dichotomous outcomes, such as our promotion/no promotion groups, yields nonnormal, unequal variance, nonlinear error terms, and predicted probabilities greater than 1.0. LR, through the procedure of maximum likelihood estimation of the log-likelihood function, avoids these statistical violations. LR yields errors that follow a binomial distribution and this feature does not affect the validity of statistical inferences (Menard, 1995). The most important assumption is that of specification error–either the wrong form of model was used or relevant predictors were excluded. Misspecification due to the wrong S-shaped model is unlikely since Hosmer and Lemeshow (1989) show the close similarity of the logistic model to alternatives. Misspecification due to omitted predictors is also unlikely since we included all predictors based on our initial selection procedures.

LR estimates the probability that a student in an observed group (e.g., retained/promoted, failed/passed) actually belongs to that group based on a set of predictors (Hosmer & Lemeshow, 2000). Although this is conceptually similar to DA, the model yields a LR coefficient (b) that can be converted into an odds ratio (exp(b)) for each predictor. This odds ratio tells us how the odds of being in the "target" group increase or decrease based on changes in the predictor values (Hosmer & Lemeshow, 2000). The data are typically normalized, transformed and scaled in order to account for the different types of predictors.

Decision tree model

DT are graphical representations of statistical predictions for how best to split a target variable into distinct categories based on a given set of predictors (James et al., 2021, p. 327). In our case, the split is based on whether the student was promoted, demoted, or remained at the current level. The DT is constructed from a set of predictors that bifurcate the sample; these bifurcation points on the tree are called nodes. A connection from the topmost node to a terminal node is called path, and each sub-sample created by the binary splits has its own path. These splits are created using the Gini index (James et al., 2021; Khichane, 2021). Iterating through every predictor, the Gini index is computed and compared; the lower it is, the lower the heterogeneity, thus the split is made by the variable with the lowest Gini index. Since a DT is a non-parametric way to differentiate between groups, there are no assumptions placed on either the raw data (the data may be continuous or categorical) or residuals, e.g., normal, equal variance, or independent (Breiman et al., 1984).

Since the visualization of a tree shows a predictor variable's cut-off value for creating the next nodes, no transformation of the original data was applied. This makes the resulting tree paths easier to trace and explain. Additionally, the trees were pruned (a compressing technique to reduce overfitting) by combining the nodes that led to the smallest increase in misclassification errors (Witten, 2011).

RESULTS

DA, LR, and DT results may be evaluated based on a true versus predicted values classification matrix (or confusion matrix) generated at the completion of the analysis cycles. This matrix shows a cross-tabulation of the original group classifications and the models' group classification predictions. We compute a variety of percentages as well as provide the actual-to-predicted classification numbers. In addition, we provide sensitivity (how well did we predict the primary group of interest) and specificity (how well did we predict the secondary group).

In general, cut-points for assigning a case to one group or another are typically varied through trial and error to yield false-positive (FP) and false-negative (FN) estimates which highlight the seriousness of each type of error. This means the best solution is not always to equate these errors because the relative costs of misclassification of a student depends on the situation and these "considerations tend to be highly subjective" (Overall & Klett, 1972, p. 248). For example, what is the cost of misclassifying a student as "promote" when they should have been held back? Conversely, what is the cost associated with classifying a student as "hold back" when they should have been promoted?

Numerous methods exist for adjusting the cut-points based on different characteristics of the data and the costs of misclassification. These typically consider unequal sample sizes, unequal dispersions, and estimates of costs of misclassifications and they do generate different classifications. The default approach is usually to assign a case to the group they have the

Table 3. Predictor importance for accelerated level promotion models

No	Predictor	Importance DT	Coefficient LR	Coefficient DA
0	avg_attempts_per_hw_problem	0.12	-0.015	0.223
1	average_days_before_first_attempt	0.14	-0.707	-0.6660
2	fraction_hws_uploaded_out_of_opened_hws	0.11	0.195	0.170
3	average_time_spent_on_hw_assignment	0.08	-1.121	-1.246
4	average_number_revisits_per_unsolved_hw_problem	0.15	-0.660	-0.730
5	avg_number_HW_revisits	0.04	-0.292	-0.207
6	fraction_lessons_attended_excluding_first	0.02	0.131	0.069
7	participation	0.25	2.286	3.023
8	behavior	0.09	1.192	1.835

Table 4. Predictor importance for honors level promotion models

No	Predictor	Importance DT	Coefficient LR	Coefficient DA
0	avg_attempts_per_hw_problem	0.05	0.165	2.4104
1	average_days_before_first_attempt	0.09	-0.062	0.029
2	fraction_hws_uploaded_out_of_opened_hws	0.02	0.160	0.134
3	average_time_spent_on_hw_assignment	0.12	-1.275	-2.142
4	average_number_revisits_per_unsolved_hw_problem	0.12	-0.491	-0.959
5	avg_number_HW_revisits	0.19	-0.229	0.030
6	fraction_lessons_attended_excluding_first	0.12	0.229	0.103
7	participation	0.21	2.905	4.295
8	behavior	0.08	1.716	4.854

highest probability of membership in. This is a maximum-likelihood procedure (Overall & Klett, 1972) and our classification matrices were based on this approach. Discriminant analyses and LRs used a 0.5 probability of target group membership criterion. The DT were pruned using a cost-complexity alpha value restriction; all classes had an equal weight of 1 and the cut off probability of target group membership (leaf nodes) was 0.5.

Predictor Importance

We present the comparison of the predictor-outcome relationships first followed by the accuracy evaluation of the models. The standardized DA and LR coefficients and their odd ratios, and DT predictor importance values for grade 4_1: accelerated students are presented in **Table 3**. The larger the standardized coefficient and the higher the DT importance value, the greater the influence of the predictor in differentiating between being retained at grade 5_1: accelerated or promoted to grade 5_2: advanced. Positive DA and LR coefficients reflect an increase in their probability of being promoted to advanced. Negative coefficients are associated with students being retained at accelerated. Odds ratios greater than 1.0 indicate how the probability of being promoted to advanced increases as the predictor values increase. Conversely, odds ratios less than 1.0 indicate how these probabilities of promotion drop for these students.

For the accelerated students the squared canonical correlation (r_c^2) = 0.15 and Wilks' lambda = 0.85 ($F = 15.0$, $p < .001$). The overall solution is highly statistically significant and the combined set of predictors accounts for 15% of the total variance in the accelerated groups. Box's (1949) M was significant ($M = 271.6$, $x^2 = 266.4$, $p < .001$). Based on an inspection of the covariance matrix determinants (a generalized measure of multivariate variance (Green, 1976)) and the sample variances, the dispersion of the promoted students was less than the retained students on eight of the nine predictors. As a group, grade 4_1: accelerated students promoted to grade 5_2: advanced showed more consistency in their online behaviors and teacher ratings than those students retained at the accelerated level.

The standardized DA and LR coefficients and their odd ratios, and DT predictor importance values for grade 4_3: honors students are presented in **Table 4**. The larger the standardized coefficient and higher the DT importance value, the greater the influence of the predictor in differentiating between being demoted to grade 5_2: advanced or retained at the grade 5_3: honors level. Positive coefficients reflect an increase in their probability of being retained at the honors level. Negative coefficients are associated with students being demoted down one curriculum level to advanced. Odds ratios greater than 1.0 indicate how the probability of being retained at the honors level increases as the predictor values increase. Conversely, odds ratios less than 1.0 indicate how these probabilities drop.

For the honors students the squared canonical correlation (r_c^2) = 0.22 and Wilks' lambda = 0.77 ($F = 10.4$, $p < .001$). This solution is also highly statistically significant and the combined set of predictors accounts for 22% of the total variance in the honors groups. Box's (1949) M was significant ($M = 95.3$, $x^2 = 91.3$, $p < .001$). Based on the covariance matrix determinants and the sample variances, the dispersion of the retained honors students was less than the demoted students on five of the nine predictors. At the grade 4_3: honors level, in contrast to the grade 4_1: accelerated level, there was less variation in the online behaviors and teacher ratings between those students retained at the grade 5_3: honors level than those who were demoted.

An inspection of the predictor Tol and VIF values (not reported) for both solutions suggests that the teacher ratings on student behavior and participation can be understood as a function of the combined relationship of the student-related predictors--a finding that is reasonable since teachers are aware of their students' study habits.

Based on **Table 3** and **Table 4**, the following tendencies across the three models were observed for both grade 4_1: accelerated and grade 4_3: honors. *Participation* has the largest importance value for both levels because those students who are active in class are usually the ones who not only pay attention and do their work, but also engage with the class, ask questions on topics they aren't sure about, and form closer relationships with the teacher, which is true for every level, regardless of its complexity. The positive coefficients suggest that as *participation* increases, the chance of the student being promoted from grade 4_1 to grade 5_2 and retained at grade 5_3 from grade 4_3 increases. *Behavior*, too, is a strong positive predictor (particularly for grade 4_3: honors students) because good behavior means you listen, pay attention, and do what you are supposed to do, which contributes to more effective learning. These two predictors, however, are not independent, as the behavior and participation assessment is performed by the same teacher who later makes the placements. We try to minimize this potential for bias by averaging the two behavior and two participation scores from the two semesters.

The large negative coefficients for *average_time_spent_on_hw_assignment* and somewhat weaker negative coefficients for *avg_number_HW_revisits* suggest that the more time some students spend on homework assignments and the more sessions they take for their homework, the less is their chance of promotion from 4_1 to 5_2 and of being retained at 5_3 from 4_3. The negative coefficients for *average_number_revisists_per_unsolved_hw_problem,* too, suggest that as the number of revisits to an unsolved homework problem increases, the chance of getting promoted from 4_1 to 5_2 and retained at 5_3 from 4_3 decreases. While trying to attempt problems that were too hard at first shows persistence, having too many revisits shows that a student has not mastered the material.

Although *fraction_hws_uploaded_out_of_opened_hws* has little bearing on the DT splits, the positive coefficients suggest that as the fraction of homework assignments uploaded increases, the chance of the student being promoted from 4_1 to 5_2 and retained at 5_3 from 4_3 increases. The effort of uploading at least a portion of a homework assignment seems to be valued by a teacher. Finally, the positive coefficients for *fraction_lessons_attended_excluding_first* suggest that students with better attendance tend to have a higher chance of being promoted from 4_1 to 5_2 and retained at 5_3 from 4_3.

Despite many similarities, successful growth on different levels does draw on different learning strategies. For example, *avg_attempts_per_hw_problem* is a much stronger predictor for grade 4_3: honors than for grade 4_1: accelerated. Although unexpected we believe this is because honors students are likely to use their attempts more frugally than accelerated students. In other words, in honors students don't resort to the 'guess and check' approach anymore.

In contrast, *average days before the first attempt* carries much less weight as a predictor for grade 4_3: honors students than for the grade 4_1: accelerated students. honors students, as a group, tend to start their homework right after class or on the day when class starts. Accelerated students, however, show much greater variability when they start their homework and those who take longer have a decreased probability of promotion to grade 5_2: advanced.

Classification Results

Although the patterns of relationships between the predictors and the binary outcomes for grade 4_1 and grade 4_3 are useful for pedagogical purposes (e.g., highlighting to students and parents the importance of attending to the homework problems soon after the lesson is completed), the ultimate criterion in our proof-of-concept investigation is the capability of each of the three models to accurately predict grade 5 curriculum placements.

Numerous authors across diverse disciplines illustrate how the interpretation of classification metrics such as sensitivity (true positive rate), specificity (true negative rate), FPs (complement of specificity), and FNs (complement of sensitivity) must be guided by the specific context in which the predictive model is applied. Parikh et al. (2008) emphasize that the relevance of these metrics is context-dependent. Hand (2009), too, argues that sensitivity/specificity trade-offs must be interpreted within domain-specific cost frameworks, not in isolation.

For example, in decisions about student promotion or retention, Baker and Inventado (2011) emphasize that high sensitivity is critical where under-identifying students at risk (i.e., FNs) can lead to lack of support for students who need it most. This means that predicting, say, that a grade 4_1: accelerated student will be retained at grade 5_1: accelerated when they could succeed if promoted to grade 5_2: advanced is harmful–low FNs

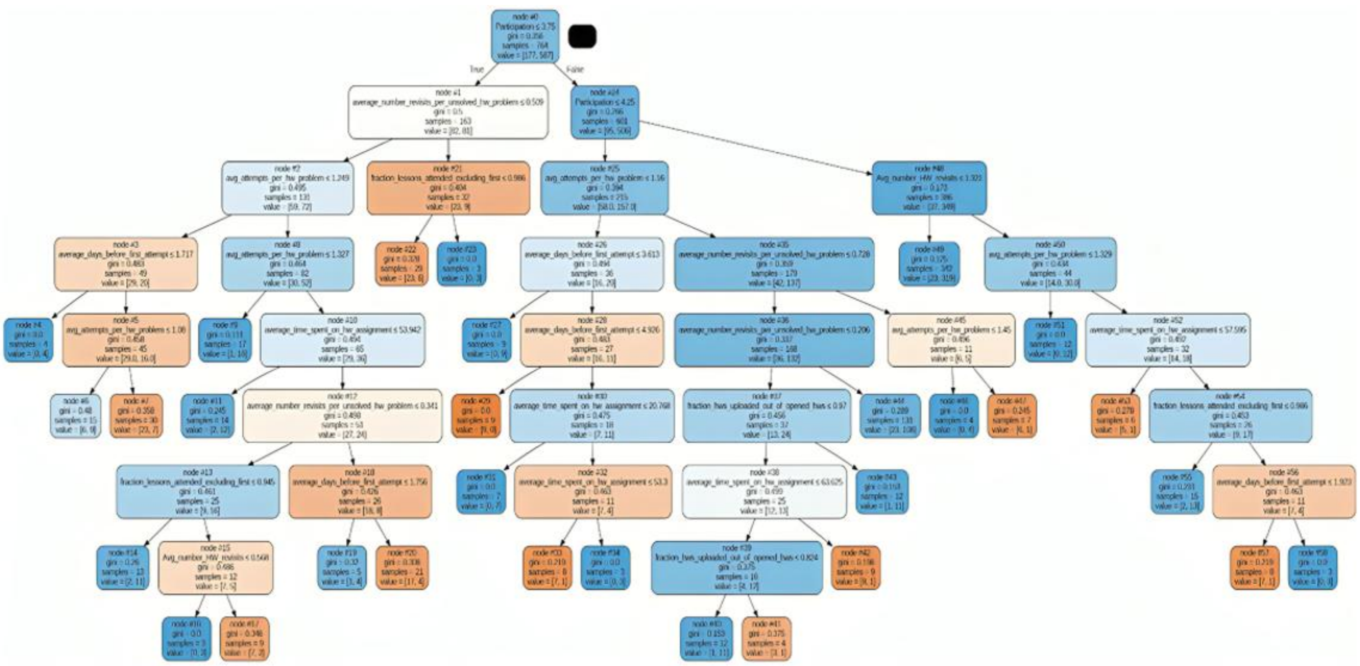


Figure 1. Decision tree for grade 4_1 promotion decisions to grade 5 (Source: Authors' own elaboration)

along with high sensitivity are preferred. Jimerson (2001), for example, illustrates how retaining students often leads to negative academic and social outcomes, so FNs (i.e., unnecessary retention) can have serious consequences.

Conversely, high specificity is important when it is critical to avoid promoting students who are not adequately prepared (Zhou et al., 2011). Here FPs can have a cost–in this case, promoting unprepared students who may struggle later. This means that while high sensitivity is good, specificity should not be ignored–a balance is often needed based on the relative cost of misclassification types. Thus, evaluating a model's performance requires not just statistical analysis but also an understanding of the practical consequences of classification errors in educational settings.

Acceptable ranges for sensitivity, specificity, FP, and FN rates in educational predictive models depend on the specific context and the consequences of misclassification. While there isn't a universal standard, several sources provide general guidelines that can be adapted to educational settings (BMJ Best Practice, n. d.; Ratliff, 2022). Sensitivity: aim for ≥ 0.80 to ensure that students who are ready for advancement are correctly identified. Specificity: aim for ≥ 0.80 to ensure that students not ready for advancement are not incorrectly promoted. Values for both are "acceptable" from 0.80 to 0.90, while values ≥ 0.90 to 1.00 are "excellent" (Ratliff, 2022). FPs: keep ≤ 0.20 to minimize the number of students incorrectly identified as ready for advancement. FNs: keep ≤ 0.20 to minimize the number of students incorrectly identified as not ready for advancement. Since FPs are the complement of specificity and FNs are the complement of sensitivity, values for both from 0.20 to 0.10 are acceptable, and ≤ 0.10 are excellent

Decisions tree results

DT were created for grade 4_1: accelerated and grade 4_3: honors. The modeled outcome for grade 4_1 was whether the student was retained in grade_5_1 or promoted to grade 5_2. For grade 4_3 the modeled outcome was whether the student was demoted in level to grade 5_2 or retained in level in grade 5_3. Each tree was constructed with one predictor at a time. Initial predictor choices were based on our classroom experiences about the student characteristics that teachers either explicitly or tacitly use when they make decisions about what level to place students in when they move up in grade–in our case from grade 4 to grade 5.

As the tree models increased in complexity with each additional predictor added to the models, the differing influences of the predictors, or lack thereof, quickly became apparent. This fluid shifting in predictor-to-outcome relationships led to numerous variations and iterations in tree construction. We present here the final two trees (grade 4_1 to grade 5, and grade 4_3 to grade 5), respectively.

Figure 1 contains the grade 4_1: accelerated predictions to either grade 5_1: accelerated (they stayed at the same level) or grade 5_2: advanced (they were promoted to the next higher level). At node #0 there are n = 177 students ("samples") who were retained in grade 5_1, there are n = 577 who were promoted to grade 5_2. The first and most influential predictor to split them further is participation with a cut-score of < = 3.75. The "true" arrow refers to paths taken by the retained

Table 5. DT classification table for grade 4_1: Accelerated

	Predicted 5_1 accelerated students: Retained	Predicted 5_2 advanced students: Promoted	Total
True 5_1 accelerated students: Retained	115 Specificity: 115/177 = 0.640	62 FPs: 62/177 = 0.350	177
True 5_2 advanced students: Promoted	25 FNs: 25/587 = 0.040	562 Sensitivity: 562/587 = 0.960	587
Total	140	624	764

Note. Overall accuracy = (115 + 562)/764 = 0.890

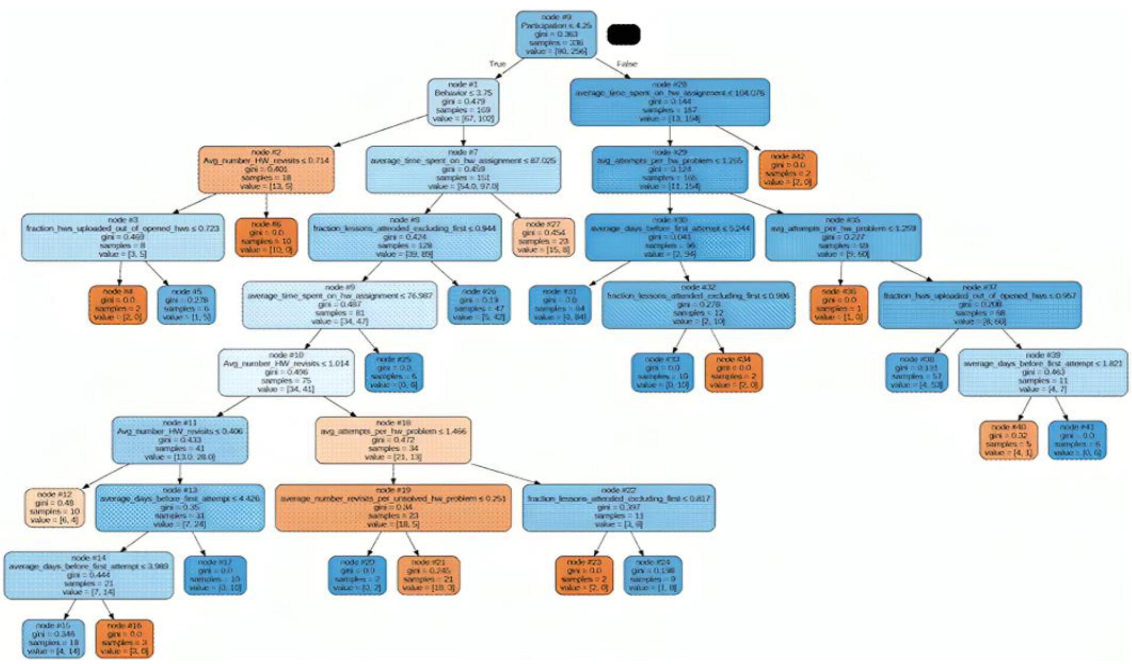


Figure 2. Decision tree for grade 4_3 promotion decisions to grade 5 (Source: Authors' own elaboration)

students and the "false" arrow shows the paths for those promoted. Each additional predictor splits these two primary outcomes into smaller and smaller sub-samples until there are no further predictors capable of differentiating between the students in a sub-sample.

Node #4 is the terminal point capturing the path that describes the performance characteristics of four grade 4_1 students who were retained at but incorrectly predicted to be at grade 5_1. Their retention path included values of participation < = 4.25, average_ number_ revisits_per_unsolved_hw_problem < = 0.51, avg_attempts_per_hw_problem < = 1.249, and average_ days_before_first_attempt < = 1.717.

The same path analysis applies for those students in grade 4_1 who were promoted to grade 5_2. Moving down the rightmost nodes and paths, we see three students at node #58 who were promoted and correctly predicted to be at grade 5_2. Their path to promotion included values of participation > 4.25, avg_number_ HW_revisits > 1.921, avg_attempts_per_hw_problem > 1.329, average_time_spent_on_hw_assignment > 57.595, fraction_lessons_attended _excluding_first > 0.986 and average_days_before_first_attempt > 1.923. Finally, we point out the numerous other paths capturing the underlying characteristics of different sets of students who were either retained or promoted. The overall accuracy of identifying grade 4_1: accelerated to grade 5 level decisions was 0.89.

Our modeling strategy was to employ all the predictors for each model. This enables us to compare the strengths of the various coefficients but more importantly it enables us to build classification matrices based on the same complete data set. For example, Table 5 presents the classification matrix for the grade 4_1: accelerated predictions for retaining students in grade 5_1: accelerated or for promoting them to grade 5_2: advanced. Sensitivity (correctly predicting promotions = 0.96) and specificity (correctly predicting retaining = 0.64) are excellent and acceptable, respectively. The FPs (incorrectly predicting promotions = 0.35) are weak while the FNs (incorrectly predicting retaining = 0.04) are excellent.

Similar to Figure 1, Figure 2 shows DT for grade 4_3: honors promotion decisions to grade 5. Here the decisions are either to be demoted to grade 5_2: advanced or retain at the grade 5_3: honors level. Of the 336 grade 4_1: honors students, 80 were demoted to grade 5_2 while 256 were retained at grade 5_3. Node #0 shows this initial split into the two groups. If their participation score is "true" < = 4.25, then we see the demoted to grade 5_2 paths; if their score is > 4.25 (the "false" path), we see them retained at grade 5_3 paths.

Table 6. DT classification table for grade 4_3: Honors

	Predicted 5_2 advanced students: Demoted	Predicted 5_3 honors students: Retained	Total
True 5_2 advanced students: Demoted	65 Specificity: 65/80 = 0.810	15 FPs: 15/80 = 0.190	80
True 5_3 honors students: Retained	16 FNs: 16/256 = 0.060	240 Sensitivity: 240/256 = 0.940	256
Total	81	255	336

Note. Overall accuracy = (65 + 240)/336 = 0.907

Table 7. LR classification table for grade 4_1: Accelerated

	Predicted 5_1 accelerated students: Retained	Predicted 5_2 advanced students: Promoted	Total
True 5_1 accelerated students: Retained	35 Specificity: 35/177 = 0.199	142 FPs: 142/177 = 0.830	177
True 5_2 advanced students: Promoted	15 FNs: 15/587 = 0.026	572 Sensitivity: 572/587 = 0.970	587
Total	50	714	764

Note. Overall accuracy = (35 + 572)/764 = 0.790

Table 8. LR classification table for grade 4_3: Honors

	Predicted 5_2 advanced students: Demoted	Predicted 5_3 honors students: Retained	Total
True 5_2 advanced students: Demoted	13 Specificity: 13/80 = 0.160	67 FPs: 67/80 = 0.840	80
True 5_3 honors students: Retained	6 FNs: 6/256 = 0.023	250 Sensitivity: 250/256 = 0.970	256
Total	18	317	336

Note. Overall accuracy = (13 + 250)/764 = 0.780

Table 9. DA classification table for grade 4_1: Accelerated

	Predicted 5_1 accelerated students: Retained	Predicted 5_2 advanced students: Promoted	Total
True 5_1 accelerated students: Retained	85 Specificity: 85/177 = 0.480	92 FPs: 92/177 = 0.520	177
True 5_2 advanced students: Promoted	77 FNs: 77/587 = 0.130	510 Sensitivity: 566/587 = 0.870	587
Total	162	602	764

Note. Overall accuracy = (85 + 510)/764 = 0.780

Although the predictors are generally the same, we see the introduction of the *behavior* variable into the model at node #1. Unlike the grade 4_1 students, the classroom behavior of the grade 4_3 students is related to their grade 5 level placement. The overall accuracy of this model is 0.91.

Table 6 presents the classification matrix for the grade 4_3: honors predictions for demoting the students down one level to grade 5_2: advanced or retaining them at the grade 5_3: honors level. Sensitivity (correctly predicting retaining = 0.94) and specificity (correctly predicting demotion = 0.81) are excellent and acceptable, respectively. The FPs (incorrectly predicting retaining = 0.19) and FNs (incorrectly predicting demotion = 0.06) are acceptable and excellent, respectively.

Logistic regression results

Table 7 presents the classification matrix for the grade 4_1: accelerated predictions for retaining at grade 5_1: accelerated or promoted to grade 5_2: advanced. Sensitivity (correctly predicting promotions = 0.97) is excellent while specificity (correctly predicting retaining = 0.199) is poor. The FPs (incorrectly predicting promotions = 0.83) are poor while the FNs (incorrectly predicting retaining = 0.04) are excellent.

Table 8 presents the classification matrix for the grade 4_3: honors predictions for being demoted to grade 5_2: advanced or retained in grade 5_3: honors. Sensitivity (correctly predicting retaining = 0.97) is excellent while specificity (correctly predicting demotion = 0.16) is poor. The FPs (incorrectly predicting retaining = 0.84) are poor while FNs (incorrectly predicting demotion = 0.023) are excellent.

Discriminant analysis results

Table 9 presents the classification matrix for the grade 4_1: accelerated predictions for retaining at grade 5_1: accelerated or promoted to grade 5_2: advanced. Sensitivity (correctly predicting promotions = 0.87) is excellent while specificity (correctly predicting retaining = 0.48) is weaker. The FPs (incorrectly predicting

Table 10. DA classification table for grade 4_3: Honors

	Predicted 5_2 advanced students: Demoted	Predicted 5_3 honors students: Retained	Total
True 5_2 advanced students: Demoted	53 Specificity: 53/80 = 0.660	27 FPs: 27/80 = 0.340	80
True 5_3 honors students: Retained	52 FNs: 52/256 = 0.020	204 Sensitivity: 204/256 = 0.800	256
Total	105	231	336

Note. Overall accuracy = (53 + 204)/336 = 0.760

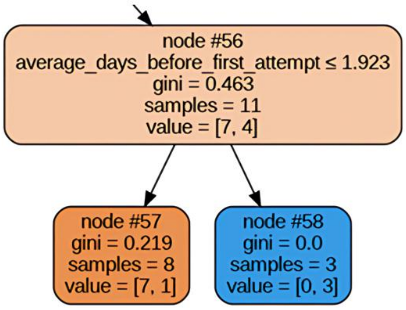


Figure 3. FN prediction (Source: Authors' own elaboration)

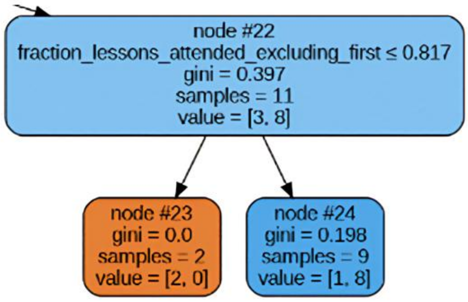


Figure 4. FP prediction (Source: Authors' own elaboration)

promotions = 0.52) are poor while the FNs (incorrectly predicting retaining = 0.13) are excellent.

Table 10 presents the classification matrix for the grade 4_3: honors predictions for being demoted to grade 5_2: advanced or retained in grade 5_3: honors. Sensitivity (correctly predicting retaining = 0.8) is acceptable while specificity (correctly predicting demotion = 0.66) is weaker. The FPs (incorrectly predicting retaining = 0.34) are weak while FNs (incorrectly predicting demotion = 0.2) are acceptable.

The DT accuracy estimates for grade 4_1: accelerated and grade 4_3: honors were 0.89 and 0.91, respectively. These are improvements over the corresponding LR (0.79 and 0.78) and DA (0.79, 0.80) estimates. The models differ, more significantly, in their FP and FN rates. The DT FP and FN rates for grade 4_1 were 0.35 and 0.04. These contrast sharply with the corresponding LR rates of 0.83 and 0.026 and the DA rates of 0.77 and 0.035. Likewise, for grade 4_3 students, the DT rates were 0.19 and 0.06 versus the corresponding rates of 0.84 and 0.023 for LR and 0.66 and 0.035 for DA. The next section explores these findings in more depth.

Misprediction Analysis

It is a truism that no statistical model will fully account for all the variation in the data. This means there is always residual variation remaining after model fitting. This residual variation may be random or may contain systematic variation worth investigating. In the present situation our three models predicted what the student's curriculum placement level in grade 5 should be as a function of a variety of student characteristics. Since we know the actual placements, we tallied not only the correct identifications but also the incorrect ones. The incorrect predicted placements, or misclassifications, can be identified and patterns and influences can be sought for why

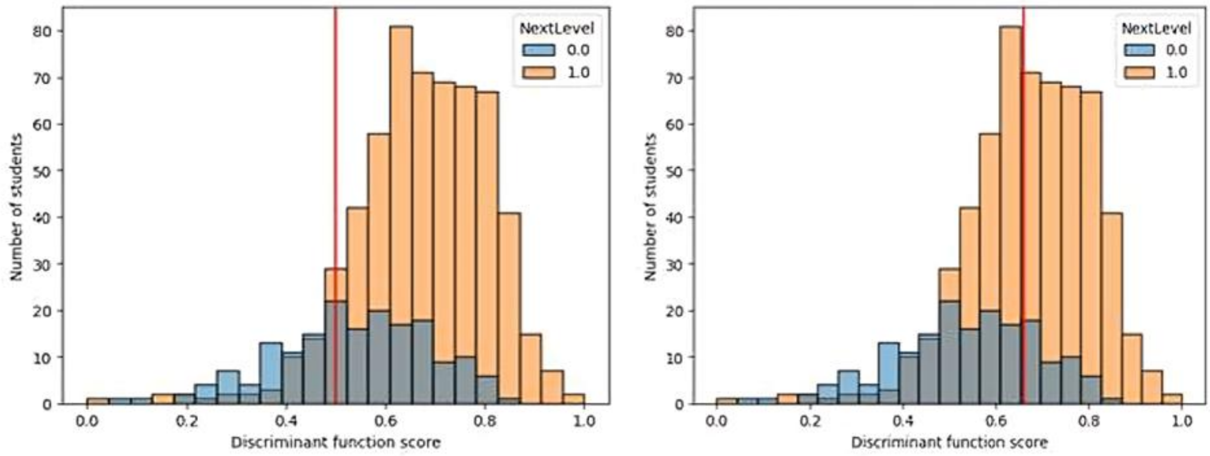
(a) some grade 4_1: accelerated students were unexpectedly promoted to grade 5_2: advanced (FN) while why some accelerated students were unexpectedly retained in the grade 5_1: accelerated level (FP) and

(b) some grade 4_3: honors students were unexpectedly retained in the honors level (FN) while some honors students were unexpectedly demoted to advanced (FPs).

Are there explanations from the teachers that shed some light on what influenced some of these FP and FN promotion decisions?

Each of our three models produced a classification matrix showing the number of known students in each promotion condition and the model's predicted number of students for that condition. Significantly, the mispredictions in the three models yield approximately the same list of misplaced students. Since there are two types of mispredictions that occur in these analyses we present a brief example of the consistency of the models. Starting with the DT analysis, we see in **Figure 3** that node #56 splits 11 students into two groups. In node #57 we see the 7 students from node #56 who were not promoted but we also see one student who was–based on the criterion in node #56 this student should be in node #58 and is an FN. Their probability of being retained in grade 5_1: accelerated level was 0.79 and 0.81 for the LR and DA analyses, respectively.

Conversely, we see in **Figure 4**, node #24, that a grade 4_3: honors student fulfilled the criterion in node ##22 and was predicted to be retained in grade 5_3: honors but was demoted to grade 5_2: advanced instead. This FP retention prediction by the DT was supported by LR and DA probabilities of 0.61 and 0.57, respectively.



Cut-off score of 0.50 Cut-off score of 0.66

Figure 5. Changing prediction cut-off score for accelerated level students (Source: Authors' own elaboration)

We mentioned earlier that all three of these models have cut-points that maximally differentiate between any two groups and that these cut-points may be altered by the researcher depending on the objectives of the study. An advantage of the DA model is the particularly easy opportunity to adjust the probability cut-point. The default cut-point probability for group assignments is 0.5, which generally aims to reduce all mispredictions, with no preference for either maximizing or minimizing FPs or FNs. In our real-life situation, however, the two types of mispredictions represent two different groups of students. Using the grade 4_1: accelerated students in **Table 9** as an example, the FN students were predicted to stay in accelerated but were, instead, promoted to advanced (3.5% out of 587 promotions). The FP students were predicted to be promoted to advanced but did not (23% out of 177 not promoted). These two types of mispredictions (and surprises) are not, in our minds, equal from the point of view of an educational process. What happens when the cut-point is changed?

For example, if the cut-point is raised to, say, 0.66, then for those 587 students who were promoted sensitivity drops from 96% to 89% while the change in FN goes from 3.5% to 11%. Fewer students will be correctly predicted for promotion (sensitivity) but more students will be promoted who were predicted to be held back (FN)–perhaps a pleasant surprise for them. For the 177 students who were not promoted, specificity increases from 23% to 42% while FPs drop from 77% to 58%. More students will be correctly predicted as being held back (specificity) with fewer predicted to be promoted but not (FP)–reducing the number of unpleasant surprises. Likewise, if the cut-point is reduced, the percentages will move in the opposite directions.

These contrasting decisions and consequences are clearly represented in **Figure 5**. These plots represent the 0.5 default strategy (left part in **Figure 5**) and 0.66 modified strategy (right part in **Figure 5**) for setting cut-points for predicting the successful identification of those grade 4_1: accelerated students who were either promoted to grade 5_2: advanced (the bars to the right of the vertical lines depicting the cut-points) or retained at the accelerated level. As the cut-point changes, so too will the correct and incorrect classifications.

What strategy should an educational institution choose? That is, what are the costs and benefits associated with changing the cut-points in ways that maximize or minimize the four cells in a confusion matrix? The answer, in the end, is not based on statistics but institutional objectives, mission and values. For example, if early intervention for remedial learning activities designed to identify and help more students at risk of not being promoted is seen as a priority policy objective, then increasing specificity by raising the cut-point is a reasonable strategy.

Our attention turns now to trying to better understand why the misclassifications occurred. We consulted the mispredicted students' attendance and classroom records, and their teachers' comments about what caused the students to be promoted/not promoted (for grade 4_1) or retained/demoted (for grade 4_3). Based on these comments (the most illustrative are provided), the following generalizations are offered.

For *accelerated* level students, a teacher generally left an otherwise successful student in accelerated without promoting for one of the following reasons:

- previous math skill gaps (results of low quiz scores): "struggles with basic math issues",
- inability to focus and behave: "easily distracted but good", and
- inability to work independently: "no independent work".

In contrast, a teacher might promote an apparently weaker student for the following reasons:

- good classwork performance even with poor homework performance: "gaps in HW, but smart",
- a student is shy and quiet, but "works best independently", and
- a student is generally very promising: "active and diligent", "very sharp".

For *honors* level students, a teacher may demote an otherwise successful student for the following:

- a student may be visibly uncomfortable with the honors level pace/ workload: "He was in my 4_3 but he was not a good match for 5_3",
- a student may be too talkative, messy, or unfocused in a fast-paced class: "Sweet child but has focus problems", and
- a student may be good at homework but very poor in class (meaning the parents help too much): "doing well but not always stable, especially for quizzes, tests and practice".

In contrast, a teacher may assign an honors level placement to a student who quantitatively might not merit it because of:

- good classwork performance even with inconsistent homework effort: "brilliant, but inconsistent" and
- some gaps in knowledge exist for an overall promising student but the teacher "vouches" for the child. In this situation (not uncommon) the teacher assigns the student to an honors class but also signs the child up for a remedial summer class: "Can try 5 honors prep and go from there".

These examples clearly illustrate that no statistical model will fully consider all the various factors that classroom teachers use when they make their curriculum promotion decisions. The models may reveal trends in the placements and individually strong influences based on quantitative performance data but the experience and reasoning of a teacher may introduce an unpredictable factor that runs counter to the otherwise generally accurate predictions.

Model Testing

Any inferential statistic based on a sample tends to have an upward bias because of idiosyncratic sampling error. In the present classification context this capitalization on chance variation means the percentages reported above, which were based on the same students used in computing the functions, are likely to shrink somewhat when the equations are applied to an independent sample (Pedhazur, 1982). A cross-validation using our original data as a "training sample" to generate the equations applied to a "calibration sample" is one way of assessing the expected shrinkage and validity of the discrimination and classification procedures. Hence, the final step in our proof-of-concept promotion prediction modeling was to cross-validate our models on an independent set of data.

Based on the 2020-2023 process data results, we tested the predictors and models on similar data from the 2023-2024 academic year. There was, however, a significant change in curriculum for the honors level, which rendered the honors results incomparable. The accelerated curriculum stayed the same and served as data for the cross-validation testing. Furthermore, since the purpose of this step is simply to see how well such a modeling process works as a precursor to actually adopting this procedure in future RSM actual operations, we present only the DT results (the LR and DA were essentially the same and offer no additional insight into the accuracy of the cross-validation analysis).

The process data consisted of 481 student records after removing all outliers and cleaning the data. **Table 11** contains the DT cross-validation classification results for the 2023-2024 data.

The cross-validation classification results in **Table 11** compared to the original results in **Table 5** show that sensitivity, specificity, and overall accuracy dropped (0.96 to 0.81 0.64 to 0.28, 0.89 to 0.73, respectively). Correspondingly, FPs and FNs increased (0.35 to 0.72 and 0.04 to 0.19, respectively). These findings are not unexpected because as the well-documented negative educational effects of COVID-19 continue to slowly decrease over time, we see students who slowly regain their former learning momentum. For example, in 2020-2023 the percentage of students staying in grade 4_1 to grade 5_1 was 43% but in 2023-2024 this number decreased to 14%. Since the models built on 2020-2023 data assume the same relationships among the predictors and promotion outcomes for the cross-validation testing, they return higher prediction errors because the predictor-to-outcome relationships have changed somewhat. Building more accurate models will mean a continuous program of database enlargement and model refinements (combined with qualitative teacher assessments).

DISCUSSION

Supporting and developing learning strategies and behaviors is a part of education that is crucial for academic success. By improving homework and classwork habits, students intake more of the curriculum offered in an educational institution while also establishing a foundation for lifelong learning opportunities. This focus on 'soft skill' learning becomes a valuable educational investment.

Different levels of the RSM curriculum are an explicit reflection of the 'stronger' and 'weaker' classes that are likely to exist in any educational institution. The emphasized transition between the levels serves as a beacon to parents and students alike, suggesting to them that they find a way to develop skills leading to progress and promotion. Skill development is an asset in any educational environment, because, although very few schools offer clearly outlined ‘level’ systems, students with different learning paces are differentiated in many institutions. It is also very fortunate that our accelerated level was available for testing, because accelerated is an entry level at RSM, hence the behavior of such students corresponds to the average body of schoolchildren.

By understanding differences in student levels and learning paces, RSM can build tailored curricula, focusing on different strategies and skills to meet the needs of all levels of students. However, understanding a pattern of out-of-sight behaviors, not always observable to parents and not always acknowledged by students themselves, may trigger individual growth within a student that increases their growth and potential for promotion.

Understanding such patterns is the objective of our evolving novel research program. Firstly, the proprietary RSM software enables us to capture student behavior in minute detail. The nature of these data brings a deeper understanding of how students approach their homework and how this contributes to their academic success.

Secondly, the data provided by RSM stretches over several contiguous years and is collected from many independent cohorts with different teachers all sharing the same curriculum. This feature helps minimize external validity errors based on a single year’s worth of data (for example, there might be a disruption for part of a school year impairing the learning, but if several contiguous year’s data are analyzed, generalizability errors are reduced). Additionally, testing the same variables, types of students, and predictors on successive years of data means that the statistical power of the results is stronger than that based on a single year.

Finally, researching statistically significant trends and patterns from several years’ data can suggest specific learning requirements and recommendations that should increase the pace of every individual student and ultimately make their growth potential noticeably higher.

Although our analysis focused on grade 4 students for technical reasons (availability of data, continuity of curricula, statistical stability), we believe that the results of the analysis and recommendations they suggest can be extended to upper grades too. Furthermore, the difference in recommendations for accelerated and honors level students represents the necessity of a differentiated approach for students with different learning paces and levels of involvement. Although not tested by us, we believe that these differences hold for any age of student, and following our recommendations can contribute to enhanced knowledge at different levels.

We answered our research questions in the following ways. Firstly, addressing RQ1, the set of statistically significant process data variables listed in **Table 1** was successful in predicting next-year next-level curriculum placement. This result was established through three independent, yet complementary, predictive models. Secondly, addressing RQ2, although there are many similarities in the online homework requirements for the two studied curriculum levels, there are meaningful differences in the recommendations we make for these differently paced students. The analysis of these differences is presented below. Finally, addressing RQ3, we discovered a set of misprediction tendencies that seem reasonable and useful to know; their analysis can be found previously.

Two of our key indicators, student participation and classroom behavior, are universally important for any educational level and environment. An actively engaged student is a contribution to any classroom as well as to themselves, clarifying finer points and using questions to create a well-indexed storage of new material in their mind that can be effectively used later. This confirms existing qualitative studies, such as Akpur (2021), using quantitative data. Although the self-learning Vygotskian approach used as the basis of RSM learning contributes to that process, the ability to successfully create a productive learning environment is our top priority and a recommendation for all levels of students.

Another universal indicator is not to leave a problem unsolved. This kind of consistency is supported by the fact that, for both curriculum levels, the potential for student promotion decreased as the number of average revisits to an unsolved problem increased. This builds upon studies such as Klein-Latucha and Hershkovitz (2024), which connect the tendency to return to unfinished tasks with persistence and establish the general positive directionality of this predictor to the student grade. However, demonstrating that, in relation to progress potential, this indicator actually exhibits a negative directionality and represents a novel finding of the present study. Since mastery of content knowledge is an important factor in considering whether to promote a student to a higher level, demote them, or keep them at the same level, this is an especially important factor our teachers consider in their placements.

Another important indicator for both levels is the average time spent on homework assignments, which is commonly used as a positive relationship with student outcome (Feng & Roschelle, 2016). Both our previous study (Ilina et al., In press) and our current results show that the relationship between the time a student spends on homework and positive educational outcomes is, in fact, not linear. If a student spends too little time on an unsolved or incorrect answer, that suggests that understanding is incomplete or addressed very haphazardly. If, on the other hand, a student spends over an hour on homework, that suggests a lack of concentration or, again, knowledge gaps. These situations are indicators of low pace and potential in any age and grade level of student. In our practice we introduce and reinforce homework pace and rhythm in order to encourage working without distractions.

Another positive relationship, however weak, was found between written homework completion and promotion chances; this is not a very common indicator which has not been used much in existing literature. Without this critical component of homework evaluation, it is hard for a teacher to know if a student fully understands how to solve the problem and the steps they took or if they are just guessing or using an external tool. Written homework is also closely associated with note-taking–an important skill that we emphasize. Organizing one's thoughts in a clear, concise and readable way is a skill more evident at higher levels of education, but acquiring it early adds to multi-tasking and reasoning skills that are useful at any level and age.

The least significant success indicators, the number of homework revisits and attendance, demonstrate that both for faster paced and slower paced students, being present in class without participation does not necessarily lead to success. Opening one's homework without completing anything also, predictably, does not count towards student success. This builds upon the existing studies that correlate attendance with academic performance (Lukkarinen et al., 2016). However, here we first begin to acknowledge the inevitable differences in approach that characterize beginners and fast-paced students that have not been outlined in the previous work on the subject. For any beginner, it takes more time to concentrate and begin work in an unfamiliar subject. Hence, it often takes a beginner 5-10 minutes to warm up at the beginning of each time they sit down to work on their homework. This means added time every time they revisit their homework and therefore, the less homework a beginner revisits, the quicker and better their homework is completed. The recommendation for accelerated students therefore can be to keep focused for as long as possible rather than break their homework completion into too many sessions. This relationship is not found for fast-paced continuing students. Therefore, the recommendation to try and complete homework assignments in as few sessions as possible is directed mainly at slower-paced beginner students. For faster-paced students, they are able to jump in immediately and retain content knowledge longer, so breaking their homework into chunks and solving it over time before the next lesson is no less efficient in overall time spent on homework.

Another variable that yields differing recommendations depending on skill level is the number of attempts required to solve a homework problem. Existing research (e.g., Davis et al., 2020) has primarily examined older students as a single group, without distinguishing between novice and advanced learners. As a result, the observed correlation between the number of attempts and overall performance has been uniformly positive. However, data from RSM offer an opportunity to expand this line of inquiry by introducing two distinct focus groups, potentially revealing more nuanced, multidimensional outcomes. Many attempts for our accelerated students relate to a higher chance of success as an indicator of persistence. However, for our honors students, too many attempts indicate the path to failure, because an incorrect answer on the first try usually indicates a bad understanding of the material or inability to detect a mistake. Our recommendation for beginner students, in general, stemming from this indicator can be to try again directly after a failure and check for calculation mistakes. For continuous and fast-paced students, however, the recommendation is different: in order to increase their potential, a student should check their work and calculations after every step, without rushing, to avoid mistakes whatsoever. This kind of clean work can also decrease the time required to complete homework assignments.

The timing of when a student begins their homework also appears to be linked to the overall potential of a beginner learner. Surprisingly, the underlying causes of delayed homework initiation–whether due to procrastination, competing commitments, or other factors–remain insufficiently explored. Nevertheless, RSM data can be leveraged to investigate not if not specifically the causes, but the outcomes of such delays. A student who begins their work on the first or second day after class seems to have more dedication than the one who does their homework at the last-minute before the next class period. Another factor reinforcing this relationship is the active memory span of an average child. That is, if a student encounters a difficult problem with homework and has to look into classwork notes to find a similar problem, they are more likely to find it and remember the strategy used in class, if that class happened a few days ago as opposed to a week ago. However, a more accomplished continuing student is likely to retain the knowledge from the previous lesson no matter how many days have passed. The memory span for RSM students is explicitly reinforced since at the beginning of every class teachers usually offer a quiz with the key problems of their homework, expecting students to remember the solution and reproduce it.

Although statistical models never account for all the variation in a set of data, our misprediction results are quite sensible. That is, the misprediction analysis suggests multiple factors beyond objective student quantitative data that our teachers may be using when they place a student differently than the statistical models predicted. Some of the reasons teachers may not promote accelerated level students are: the student still demonstrates gaps in their math knowledge and skills and struggles with foundational concepts, which are not reviewed as much or at all at higher level; the student is unable to focus or behave, which would make it difficult for them to keep up at a faster pace; and the student is unable to work independently without teacher intervention and support, which decreases at higher levels where lessons are more led and centered around students than the teacher. In contrast, if students demonstrate more of these abilities of higher classwork performance, independent work, and quick-thinking, they may be promoted even if they seem somewhat weak in indicators such as homework and particular assessment performance.

Similar considerations go into decisions about placements for honors students such as whether a student is able to handle the faster pace of work, focus and keep on track, and work more independently without reliance on teacher intervention and support. Another indicator is that a student has high homework performance but lower classwork and assessment scores, which can indicate that there is heavy parent involvement or external tools at play but a student has not really mastered the content. All of these factors are considered to see if the student is a good fit for the honors level and whether it is the right environment to ensure their success in progressing in their mathematical education. Inconsistent performance can go either way depending on the teacher as some hold back or demote on this basis as it may indicate a student is unable to persist at a higher level, while others believe in their potential and try to address how to work with parents to elevate their performance on a more consistent basis.

Our recommendation to teachers is that they be aware and cognizant of the interwoven web of reasons they use for placements. They should try, in particular, to avoid introducing their personal attitude towards students in the placement process. In contrast, it may be reasonable to override an otherwise obvious placement based on the child's "character". Although we offer no guideline on how to assess "character", this evaluative factor is where a teacher's judgment is invaluable in determining a placement, as only they have been able to observe the student's behavior over the course of 36 weeks, that is in the best interests for the continuing development of a child. Our management teams also keep an eye out for outliers where placements are not supported with performance indicators and they question teachers to ensure their reasoning for promotion or demotion are sound.

In summary, we have created a project-based research course for highly motivated students who have demonstrated their skills in mathematics and calculus, in particular. The curriculum is designed to teach them the basics of programming, data analysis and statistics but also, more importantly, to build a research mindset. This includes asking relevant research questions, finding a way to answer them using available data resources and how to interpret the results and handle exceptions. These tasks are all made easier by the fact that our students have an internal student's view of RSM operations and objectives. As a curriculum addition based on the experiences of the previous cohort of students, the current cohort created a draft of an educational research conference proposal and presented their independent contributions to this proposal through a Zoom pre-conference rehearsal.

This research program allows our students to gain skills that are unique for high school students. The ultimate results of the class, i.e., publications and conference presentations, can be tangibly included in a college application. These opportunities have been widely publicized and have increased parent and student interest in our program, resulting in two research cohorts starting in the 2024-2025 academic year. Our new dual curriculum consists of parallel research experiences towards parallel goals–offering educational opportunities that provide the highest degree of growth and self-learning potential for students.

CONCLUSION

In conclusion, our research has shown that a fixed set of predictors–drawing on both student-related process data and teacher-related observational data–can be used to predict students' level placements for the next academic year at RSM, namely, promotions, retentions, and demotions. Given that promotions in the RSM system reflect a student's broader academic growth potential, our findings carry significance beyond the immediate educational context.

We found that different predictors carry different weights for new students versus continuing students with more established work habits and faster working pace. This insight allowed us to develop a set of targeted recommendations which, when implemented, increase the likelihood of placement at a higher RSM level—or at least help avoid demotion. In general education settings, these recommendations have the potential to enhance student progress potential as perceived by teachers.

Our use of aggregated longitudinal data added stability to the model coefficients, increasing the reliability of our results. Furthermore, the process data we analyzed proved valuable, offering insight into student behaviors that align with teacher assessments. Some of the most informative predictors–such as how often a student returned to previously abandoned problems–are not commonly used in similar analyses, highlighting the innovative aspect of our approach.

A promising direction for future research would be to explore which predictors perform best in each of the three models and how to define optimal cut-points that account for the costs of misclassification. By combining the three models using best-fitted predictors and optimal cut-scores, we can further reduce the rate of misprediction and refine the tools available for guiding student placement decisions.

Author contributions: MK: responsible for student work supervision and contributed to data retrieval and management; **SA:** provided overall project supervision and handled legal and organizational aspects; **MS:** focused on academic implications and conducted analytical work; **JB:** advised on research directionality and broader perspectives; **AM:** conducted qualitative analysis and advised on research directionality; **OI:** responsible for data retrieval and management; **IJ, YM, MP, SS, & VV:** participated as student researchers; & **LL:** supervised the project, contributed to the statistical methodology, and shaped the research setting. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study does not require any ethical approval. No active data collection, interaction, or intervention with human subjects was involved. It is a retrospective analysis of archival data. The dataset used is fully de-identified, with no direct or indirect identifiers that could be used to re-identify individuals. The authors further stated that the data do not include private, sensitive, or confidential information. The study adhered to the ethical principles of the American Psychological Association ethical standards.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Akpur, U. (2021). Does class participation predict academic achievement? A mixed-method study. *English Language Teaching Educational Journal, 4*(2), 148-160. https://doi.org/10.12928/eltej.v4i2.3551

Algozzine, B., Wang, C., & Violette, A. S. (2011). Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavior Interventions, 13*(1), 3-16. https://doi.org/10.1177/1098300709359084

Alhassan, A., Zafar, B., & Mueen, A. (2020). Predict students' academic performance based on their assessment grades and online activity data. *International Journal of Advanced Computer Science and Applications, 11*(4). https://doi.org/10.14569/IJACSA.2020.0110425

Alshanqiti, A., & Namoun, A. (2020). Predicting student performance and its influential factors using hybrid regression and multi-label classification. *IEEE Access, 8*, 203827-203844. https://doi.org/10.1109/ACCESS.2020.3036572

Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., & Gates, K. M. (2023). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods, 55*(6), 3026-3054. https://doi.org/10.3758/s13428-022-01939-9

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson, & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61-75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Beckman, R. J., & Cook, R. D. (1983). Outlier……….s. *Technometrics, 25*(2), 119-149. https://doi.org/10.2307/1268541

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics*. John Wiley and Sons. https://doi.org/10.1017/S0033312300004397

Bezirhan, U., von Davier, M., & Grabovsky, I. (2020). Modeling item revisit behavior: The hierarchical speed–accuracy–revisits model. *Educational and Psychological Measurement, 81*(2), 363-387. https://doi.org/10.1177/0013164420950556

BMJ Best Practice. (n.d.). *Diagnostic test studies: assessment and critical appraisal*. https://bestpractice.bmj.com/info/toolkit/learn-ebm/diagnostic-test-studies-assessment-and-critical-appraisal

Bock, D. R. (1975). *Multivariate statistical methods in behavioral research*. McGraw-Hill.

Botchway, J., Babs-Olawoyin, E., & Bakare, S. (2024). Predicting academic performance among students using machine learning (HSE case study) [MBA thesis, St. Petersburg School of Economics and Management]. https://doi.org/10.13140/RG.2.2.10296.15363

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika, 36*, 317-346. https://doi.org/10.2307/2332671

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Chapman and Hall. https://doi.org/10.1201/9781315139470

Chavez, H., Chavez-Arias, B., Contreras-Rosas, S., Alvarez-Rodríguez, J., & Raymundo, C. (2023). Artificial neural network model to predict student performance using nonpersonal information. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1106679

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis*. Wiley.

Cornell-Farrow, S., & Garrard, R. (2020). Machine learning classifiers do not improve the prediction of academic risk: Evidence from Australia. *Communications in Statistics: Case Studies, Data Analysis and Applications, 6*(2), 228-246. https://doi.org/10.1080/23737484.2020.1752849

Cruz Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon, 6*. https://doi.org/10.1016/j.heliyon.2020.e04081

Davis, M. C., Duryee, L. A., Schilling, A. H., Loar, E. A., & Hammond, H. G. (2020). Examining the impact of multiple practice quiz attempts on student exam performance. *Journal of Educators Online, 17*(2). https://www.thejeo.com/archive/2020_17_2/davis_duryee_schilling_loar__hammond

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems, 49*, 498-506. https://doi.org/10.1016/j.dss.2010.06.003

Dixon, W. J. (1953). Processing data for outliers. *Biometrics, 9*, 74-89. https://doi.org/10.2307/3001634

Doctor, A. C. (2023). A predictive model using machine learning algorithm in identifying student's probability on passing semestral course. *International Journal of Computing Sciences Research, 7*, 1830-1856. https://doi.org/10.25147/ijcsr.2017.001.1.135

Epstein, J. L. (1983). *Homework practices, achievements, and behaviors of elementary school students*. Center for Research on Elementary and Middle Schools.

Fan, H., Xu, J., Cai, Z., He, J., & Fan, X. (2017). Homework and student's achievement in math and science: A 30-year meta-analysis, 1986-2015. *Educational Research Review, 20*, 35-54. https://doi.org/10.1016/j.edurev.2016.11.003

Fateen, M., & Mine, T. (2021). Predicting student performance using teacher observation reports. In *Proceedings of the 14th International Conference on Educational Data Mining* (pp. 481-486).

Feng, M., & Roschelle, J. (2016). Predicting students' standardized test scores using online homework. In *Proceedings of the 3rd ACM Conference on Learning at Scale* (pp. 213-216). ACM. https://doi.org/10.1145/2876034.2893417

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Cordeiro Galvão van Erven, G. (2018). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research, 94*, 335-343. https://doi.org/10.1016/j.jbusres.2018.02.012

Fox, J. (1991). *Regression diagnostics*. SAGE. https://doi.org/10.4135/9781412985604

Gnanadesikan, R. (1977). *Statistical data analysis of multivariate observations*. Wiley.

Green, P. E. (1976). *Mathematical tools for applied multivariate analysis*. Academic Press.

Hair, J. F., Anderson, R. E., & Tatham, R. L. (1987). *Multivariate data analysis* (2nd ed.). Macmillan.

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning, 77*(1), 103-123. https://doi.org/10.1007/s10994-009-5119-5

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley. https://doi.org/10.1002/0471722146

Huberty, C. J. (1994). *Applied discriminant analysis*. Wiley.

Huberty, C. J., Wisenbaker, J. M., & Smith, J. C. (1987). Assessing predictive accuracy in discriminant analysis. *Multivariate Behavioral Research, 22*(3), 307-329. https://doi.org/10.1207/s15327906mbr2203_4

Ilina, O., Antonyan, S., Mirny, A., Lepsky, O., Brodskaia, J., Balsara, M., Keyhan, K., Mylnikov, A., Quackenbush, A., Tarlie, J., Tyutyunik, A., Venkatesh, R., & Ludlow, L. H. (2023). Fourth grade student mathematics performance in an after-school program before and after COVID-19: Project-based learning for teachers and students as co-researchers. *Journal of Teacher Action Research, 10*(2). https://jtar-ojs-shsu.tdl.org/JTAR/article/view/94

Ilina, O., Antonyan, S., Kosogorova, M., Mirny, A., Brodskaia, J., Singhal, M., Belakurski, P., Iyer, S., Ni, B., Shah, R., Sharma, M., & Ludlow, L. (In print). Understanding fourth-grade student achievement using process data from student's web-based/online math homework exercises. *Educational Sciences*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. https://doi.org/10.32614/CRAN.package.ISLR2

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30*(3), 420-437. https://doi.org/10.1080/02796015.2001.12086124

Karalar, H. & Gürüler, H. (2021). Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *International Journal of Educational Technology in Higher Education, 18*, Article 63. https://doi.org/10.1186/s41239-021-00300-y

Kassarnig, V., Mones, E., Bjerre-Nielsen, A., Sapiezynski, P., Dreyer Lassen, D., & Lehmann, S. (2018). Academic performance and behavioral patterns. *EPJ Data Science, 7*(10). https://doi.org/10.1140/epjds/s13688-018-0138-8

Kendall, M., & Stuart, A. (1976). *The advanced theory of statistics* (3rd. ed.). Hafner.

Khichane, M. (2021). *Le machine learning avec Python: De la théorie à la pratique* [Machine learning with Python: From theory to practice]. ENI.

King, E., Orazem, P., & Paterno, E. (2008). Promotion with and without learning: Effects on student enrollment and dropout behavior. *The World Bank Economic Review, 30*. https://doi.org/10.1093/wber/lhv049

Klecka, W. R. (1980). *Discriminant analysis*. SAGE Publications. https://doi.org/10.4135/9781412983938

Klein-Latucha, O., & Hershkovitz, A. (2024). When leaving is persisting: Studying patterns of persistence in an online game-based learning environment for mathematics. *Journal of Learning Analytics, 11*(2), 42-51. https://doi.org/10.18608/jla.2023.8219

Kruskal, W. H. (1960). Some remarks on wild observations. *Technometrics, 2*(1), 1-3. https://doi.org/10.1080/00401706.1960.10489875

Lachenbruch, P.A. (1975). *Discriminant analysis*. Hafner.

Ludlow, L. H. (1999). The structure of the job responsibilities scale: A multi-method analysis. *Educational and Psychological Measurement, 59*, 962-975. https://doi.org/10.1177/00131644995960 06

Ludlow, L. H., Fiona, E., Cochran-Smith, M., Newton, A., Trefce,r K., Klein, K., Grudnoff, L., Haigh, M., & Hill, M. (2017). Visualizing teacher education as a complex system: A nested simplex systems approach. *Complicity: An International Journal of Complexity and Education, 14*(1), 36-79. https://doi.org/10.29173/cmplct26053

Ludlow, L. H., & Klein, K. (2014). Suppressor variables: The difference between "is" versus "acting as". *Journal of Statistics Education, 22*(2), 1-28. http://www.amstat.org/publications/jse/v22n2/ludlow.pdf

Lukkarinen, A., Koivukangas, P., & Seppälä, T. (2016). Relationship between class attendance and student performance. *Procedia-Social and Behavioral Sciences, 228*, 341-347. https://doi.org/10.1016/j.sbspro.2016.07.051

Menard, S. (1995). *Applied logistic regression analysis*. Sage.

Mogessie, M., Riccardi, G., & Rochetti, M. (2015). Predicting students' final exam scores from their course activities. In *Proceedings of the 2015 IEEE Frontiers in Education Conference* (pp. 1-9). IEEE. https://doi.org/10.1109/FIE.2015.7344081

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences, 11*(1), Article 237. https://doi.org/10.3390/app11010237

Oppong, E., Shore, B., & Muis, K. (2018). Clarifying the connections among giftedness, metacognition, self-regulation, and self-regulated learning: Implications for theory and practice. *Gifted Child Quarterly, 63*, 102-115. https://doi.org/10.1177/0016986218814008

Overall, J. E., & Klett, C. J. (1972). *Applied multivariate analysis*. McGraw-Hill.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology, 56*(1), 45-50. https://doi.org/10.4103/0301-4738.39395

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). Harcourt Brace College Publishers.

Raj, V., & Vannan, M. (2020). Predicting student failure in university examination using machine learning algorithms. *International Journal of Innovative Technology and Exploring Engineering, 9*, 2278-3075. https://doi.org/10.35940/ijitee.E2643.039520

Ramdass, D., & Zimmerman, B. (2011). Developing self-regulation skills: The important role of homework. *Journal of Advanced Academics, 22*, 194-218. https://doi.org/10.1177/1932202X1102200202

Ratliff, K. (2022). How to Interpret Sensitivity and Specificity. WPS Publishing. https://www.wpspublish.com/blog/understanding-sensitivity-and-specificity-casl-2

Rosário, P., Núñez, J., Vallejo, G., Cunha, J., Nunes, T., Mourão, R., & Pinto, R. (2015). Does homework design matter? The role of homework's purpose in student mathematics achievement. *Contemporary Educational Psychology, 43*, 10-24. https://doi.org/10.1016/j.cedpsych.2015.08.001

Roschelle, J., Feng, M., Murphy, R., & Mason, C. (2016). Online mathematics homework increases student achievement. *AERA Open, 2*(4). https://doi.org/10.1177/2332858416673968

Russian School of Mathematics. (2020). *About RSM: The Russian School of Mathematics*. https://www.russianschool.com/about

Shechtman, N., Roschelle, J., Feng, M., & Singleton, C. (2019). An efficacy study of a digital core curriculum for grade 5 mathematics. *AERA Open, 5*(2). https://doi.org/10.1177/2332858419850482

Tatsuoka, M. M. (1971). *Multivariate analysis*. Wiley.

Valle, A., Regueiro, B., Núñez, J. C., Rodríguez, S., Piñeiro, I., & Rosário, P. (2016). Academic goals, student homework engagement, and academic achievement in elementary school. *Frontiers in Psychology, 7*, Article 463. https://doi.org/10.3389/fpsyg.2016.00463

Vygotsky, L. (1926). *Educational psychology*. St. Lucie Press.

Wickens, T. D. (1995). *The geometry of multivariate statistics*. Erlbaum.

Wickramasinghe, I., Aragon, R., & Valles Jr., J. R. (2024). Influential factors on elementary students' academic performance and identifying the appropriate performance group. *Discover Education,* *3*, Article 76. https://doi.org/10.1007/s44217-024-00167-x

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Elsevier. https://doi.org/10.1016/B978-0-12-374856-0.00001-8

Wu, D., Li, H., Zhu, S., Yang, H. H., Bai, J., Zhao, J., & Yang, K. (2023). Primary students' online homework completion and learning achievement. *Interactive Learning Environments, 32*(8), 4469-4483. https://doi.org/10.1080/10494820.2023.2201343

Zabriskie, C., Yang, J., Devore, S., & Stewart, J. (2019). Using machine learning to predict physics course outcomes. *Physical Review Physics Education Research, 15*, Article 020120. https://doi.org/10.1103/PhysRevPhysEducRes.15.020120

Zhou, X. H., Obuchowski, N. A., & McClish, D. K. (2011). *Statistical methods in diagnostic medicine* (2nd ed.). Wiley. https://doi.org/10.1002/9780470906514

Zimmerman, B. J., & Kitsantas, A. (2005). Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs. *Contemporary Educational Psychology, 30*, 397-417. https://doi.org/10.1016/j.cedpsych.2005.05.003

https://www.ejmste.com