

Math course sequences in grades 6–11 and math achievement in Mississippi

See <https://go.usa.gov/xyfzC> for the full report.

Appendix A. Methods

The study team used the statistical package *TraMineR* (Gabadinho, Ritschard, Müller, & Studer, 2011) to describe the student math sequences in Mississippi. *TraMineR* is designed specifically for sequence analysis and is able to both identify and group similar sequences for use in examining how sequences are related to explanatory factors, such as race/ethnicity, in correlational analyses.

Associations between math sequences and college-ready performance in math on the ACT were modeled using classification and regression tree (CART) analyses. CART is a statistical technique used to classify individuals into mutually exclusive subgroups, with the results presented in a decision tree (Breiman, Friedman, Olshen, & Stone, 1984). It does this by identifying the best predictors and predictor levels that most efficiently split the sample into the most similar subgroups of individuals who are identified as at risk or not at risk based on their observed scores. A variable may appear in the CART model multiple times because the search for the single variable that will result in the best split to the data includes all variables at each split (Therneau & Atkinson, 2013). Previous studies have found CART results to be consistent with those from logistic regression (Koon, Petscher, & Foorman, 2014) and easier to understand because of CART's graphic format.

Finally, the study team conducted a supplemental logistic regression analysis for confirmatory purposes. Sequence analyses and CART analyses were conducted in R, and the logistic regression analysis was conducted in SPSS.

Sequence analysis

The math sequences of students who started grade 6 in 2011/12 (research question 1) were analyzed using the following broad steps:

1. Each element in the sequence chain consisted of the enrolled math course, or combination of courses, in a specific year from grade 6–11. Enrollment in a single statewide course in a school year was captured using existing course numbers, and the study team created course numbers to reflect combinations of two or more statewide courses (such as Algebra I and Geometry) taken in the same year.
2. The study sample included all students enrolled in grade 11 in a Mississippi public high school during the 2016/17 school year who had an ACT math score from February or March 2017 and recorded coursework in grade 6. Data were available for 27,680 of the 32,837 eligible students; 2,259 students were missing ACT math scores in grade 11, and 2,898 students were missing coursework in grade 6 (table A1). The sample of students missing ACT math scores in grade 11 included more male students and more students with missing data on eligibility for the national school lunch program than the analytic sample, and the sample of students missing recorded coursework in grade 6 consisted of more White students and fewer students eligible for the national school lunch program than the analytic sample. Students for whom math coursework in grades 7–11 was missing were retained in the sample; missing a math course was considered meaningful and reflective of no

math coursework during the school year. The number of students in the sample with missing coursework was 410 in grade 7, 383 in grade 8, 486 in grade 9, 560 in grade 10, and 1,333 in grade 11. Of students with missing coursework in grade 11 (the grade with the largest number of students missing data), 52 percent were male, 45 percent were Black, and 70 percent were eligible for the national school lunch program.

3. Sequence objects were created by *TraMineR* (Gabadinho et al., 2011) functions. The functions store the sequences along with their attributes in several formats, allowing for different descriptive summaries to be produced, such as sequence frequency tables.
4. Optimal matching distances were computed and then used to make a typology of the sequences based on similarity of courses. Ward hierarchical clustering was specified within the statistical package *cluster* (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2018). The most frequent sequences within each cluster were used to represent the cluster.

Table A1. Characteristics of grade 6–11 students in Mississippi excluded from the study, by reason for exclusion (percent)

Student characteristic	Excluded for missing ACT math score in grade 11 (n = 2,259)	Excluded for missing coursework in grade 6 (n = 2,898)	Study sample (n = 27,680)
Gender			
Male	54.5	50.0	49.0
Race/ethnicity			
Black	52.1	32.1	50.9
White	41.1	53.8	45.2
Hispanic	3.5	6.4	2.1
Eligibility for the national school lunch program			
Eligible	62.0	50.3	67.3
Not eligible	17.3	44.4	32.0
Missing	20.7	5.2	0.7

Source: Authors' analysis of data from the Mississippi Department of Education.

Descriptive statistics

After deriving the sequence clusters, descriptive statistics were used to provide the average score on the Mississippi Curriculum Test, Second Edition, in grade 5, the average ACT math score, and the percentage meeting the ACT college readiness benchmark in math in grade 11, as well as the demographic profile (including gender, race/ethnicity, and eligibility for the national school lunch program) of each cluster (research question 2).

Classification and regression tree analysis

Variables used in the CART model to predict college-ready performance in math on the ACT (research question 3) included grade 5 math achievement, gender, race/ethnicity, eligibility for the national school lunch program, and cluster membership.

Math scores on the ACT in grade 11 were coded to indicate whether a student met the ACT college readiness benchmark in math. Scores at or above the benchmark were coded 1 for not at risk, and scores below the benchmark were coded 0 for at risk. About 19 percent of the sample met the ACT college readiness benchmark in math, and 81 percent of the sample failed to meet the benchmark.

The dataset was split into a calibration dataset (used to build the CART models), consisting of a random sample of 80 percent of the students, and a validation dataset (used to test the CART models), consisting of the remaining 20 percent. CART analyses were run using the *rpart* package (Therneau & Atkinson, 2018).

The model is represented generally by

$$\text{ACT college readiness benchmark status} \sim \text{Grade 5 math score} + \text{Gender} + \text{Race/ethnicity} \\ + \text{Eligibility for the national school lunch program} + \text{Cluster membership.}$$

The study specified a minimum split size of 100 students and tenfold cross-validation for evaluating the quality of the prediction tree and determining the appropriate minimum complexity parameter for pruning the tree (Breiman et al., 1984). CART analysis accommodates the use of a combination of continuous and categorical predictors without additional specifications.

As with other statistical methods, the principle of parsimony is applicable to CART models. This principle suggests that the simplest model that fits the data is often the best model. In a CART model this principle is applied by pruning the classification tree using model specifications so that the resulting tree is not overfit to the data. As such, a minimum reduction in the cross-validation relative error (that is, a minimum complexity parameter) was added to the model specifications in a revised model. The minimum complexity parameter specifies the minimum decrease in the overall lack of fit that must result from an additional split. The study used the default minimum complexity parameter value of .01 to prune the tree. Plots of the cross-validation relative error against minimum complexity parameter values were consulted for this decision. The pruned tree with the final classification rules are shown in figure 2 in the main text.

The classification rules were applied to the validation dataset to predict group membership and to derive the classification tables used to calculate the overall classification accuracy rate (the proportion of students who were correctly identified as either meeting or not meeting the ACT college readiness benchmark in math) and four standard measures of classification accuracy that indicate how well the analysis accurately identifies students who are at risk:

- Sensitivity: the percentage of students predicted to be at risk by the model among all students who have a math score below the ACT college readiness benchmark—or the number of true positives divided by the sum of true positives and false negatives.
- Specificity: the percentage of students predicted to be not at risk by the model among all students who have a math score at or above the ACT college readiness benchmark—or the number of true negatives divided by the sum of true negatives and false positives.
- Positive predictive power: the percentage of students who have a math score below the ACT college readiness benchmark among all students who are predicted to be at risk—or the number of true positives divided by the sum of true positives and false positives.
- Negative predictive power: the percentage of students who have a math score at or above the ACT college readiness benchmark among all students who are predicted to be not at risk by the model—or the number of true negatives divided by the sum of true negatives and false negatives.

The overall classification accuracy rate of the decision rules was 88 percent when tested with the validation sample (table A2). The sensitivity rate (95 percent) was much higher than the specificity rate (60 percent). However, the study team considered it more important to judge the prediction model based on the sensitivity rate because the intention is to identify students at risk of low performance in math on the ACT. Similarly, the positive predictive power rate exceeded the negative predictive power rate.

Table A2. Standard measures of classification accuracy for the classification and regression tree and logistic regression analysis results (percent)

Model	Sensitivity rate	Specificity rate	Positive predictive power rate	Negative predictive power rate	Overall classification accuracy rate
Classification and regression tree (n = 27,680)	95	60	91	72	88
Logistic regression (n = 25,841)	95	60	95	60	88

Source: Authors' analysis of data from the Mississippi Department of Education.

Finally, logistic regression was used to validate the CART model. Logistic regression is an extension of simple or multiple regression, whereby a dichotomously scored dependent variable is regressed on one or more independent variables. In this way, logistic regression results are products of the general formula:

$$\ln \left[\frac{p}{1-p} \right] = \beta_0 + \beta_1(X)$$

where the beta coefficients β_0 and β_1 can then be used to estimate predicted log-odds, which can be converted to a predicted probability of success on the outcome, in this case meeting the ACT college readiness benchmark in math.

The data file used in the CART analysis was also used in the logistic regression analysis. However, since data imputation was not used in the model, students with missing test scores in grade 5 and missing data on eligibility for the national school lunch program were not included in the analysis, leaving the sample at 25,841 students. This differs from the CART analysis, which can use surrogate predictors when the primary predictor is missing data. Like the CART analysis, predictors used in the modeling included grade 5 math achievement, gender, race/ethnicity, eligibility for the national school lunch program, and cluster membership. Predictors were added to the model using the forward selection option. This approach resulted in the selection of two steps based on the improvement in the Nagelkerke pseudo R^2 , which is estimated by means of maximum likelihood in logistic regression and can be interpreted in the same way as the R^2 in an ordinary least squares regression analysis. Grade 5 math achievement was added in step 1, with a Nagelkerke pseudo R^2 of .521, which increased to .558 when race/ethnicity was added to the model in step 2. Subsequent additions of cluster membership and eligibility for the national school lunch program did not improve the Nagelkerke pseudo R^2 over .02, which served as the criteria for adding additional predictors to the model. Sensitivity for this model was found to be 95 percent, while specificity was lower, at 60 percent, both of which are consistent with the CART results (see table A2).

References

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.
- Gabardinho, A., Ritschard, G., Müller, N.S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.
- Koon, S., Petscher, Y., & Foorman, B.R. (2014). *Using evidence-based decision trees instead of formulas to identify at-risk readers* (REL 2014–036). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2018). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.7-1.

Therneau, T. M., & Atkinson, E. J. (2013). *An introduction to recursive partitioning using the RPART routines* (Mayo Foundation Technical Report). Retrieved December 16, 2013, from <http://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>.

Therneau, T. M., & Atkinson, E. J. (2018). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-13. <https://CRAN.R-project.org/package=rpart>.