IES Institute of Education Sciences

APPENDIXES
June 2021

Regional Educational Laboratory Southwest
At American Institutes for Research

Identifying Indicators that Predict Postsecondary Readiness and Success in Arkansas

See https://go.usa.gov/xHukK for the full report.

Appendix A. Data and methods

This appendix describes the study data, characteristics of the study population, and analysis methods used.

Data sources

Data used to address all three research questions came from four sources:

- Arkansas Department of Education.
- Arkansas Division of Higher Education.
- National Student Clearinghouse.
- National Center for Educational Statistics Common Core of Data (U.S. Department of Education, n.d.).

Independent variables. Data from the Arkansas Department of Education was used to construct the study's independent variables for all three research questions, including student background characteristics and student status on the middle school and high school indicators. The data included in each of these files are described below.

- **Demographic records.** These data include the student's grade 6 school district, gender, race/ethnicity, national school lunch program status, English learner status, disability designation, and grade level for each academic year. The study team used demographic data exclusively from the grade 6 year. One additional variable, "older than the typical age in grade 6" (defined following Arkansas guidelines as being age 13 or older on August 1 before the student's grade 6 year), was constructed by the study team as a control for student background. This lenient threshold allows for students to enter kindergarten (and subsequent grades) a year later than the state standard without yet being coded as older than the typical age in grade 6.
- **Attendance records.** These data include information on the number of enrolled days a student was present and absent—considered for the first academic year in which students were in a particular grade. Percent of days present was calculated for each grade level with reference to total days enrolled (assessed by adding days present and days absent). A first indicator variable was created that took a value of 1 if the student was present more than 95 percent of days for all years of the grade level (in middle school or high school). A second indicator variable was created that took a value of 1 if the student was present 91–95 percent of days in at least one year and not present 90 percent or fewer days in any year. These two indicator variables left a remaining baseline (or reference group) category that indicated that a student was present 90 percent or

fewer days in at least one year (absent 10 percent of days enrolled, referred to as chronically absent). These operationalizations follow the guidance in the Arkansas Every Student Succeeds Act (ESSA) plan and establish a reference group category for regression models that adheres to the definition of chronic absence used in Arkansas and other states (Chang et al., 2019; Jordan & Miller, 2017).

- **Student assessment records.** These data include whether students demonstrated proficiency as established by the Arkansas Department of Education on state English language arts, math, and science assessments during middle school, as well as for high school math and science. State assessments of math and English language arts are administered in all three middle school grades, whereas assessments of middle school and high school science and high school math are typically administered in a particular grade or for a particular course. The specific assessments include the Arkansas Comprehensive Testing, Assessment, and Accountability Program through 2013/14 (the year on-time students in the study cohorts were in grades 10 and 11), and the Partnership for Assessment of Readiness for College and Careers through 2014/15 (the year on-time students in the study cohorts were in grades 11 and 12). In 2015/16 the Arkansas Department of Education switched to the ACT Aspire, but the two grade 6 cohorts had already moved beyond high school testing by then.
    - For middle school English language arts the constructed indicator measures whether a student scored proficient or above in grade 8; if no testing information was available for the student in grade 8 (which was the case for 5.8 percent of observations), the study team based the indicator on grade 7 information (which was available for 44 percent of those missing cases). If grade 7 information was missing, the study team used grade 6 information (which was available for an additional 30 percent of missing cases). If grade 8, 7, and 6 information was unavailable, the student was considered not to have demonstrated proficiency.
    - For middle school math the constructed indicator measures whether a student scored proficient or above in grade 8; if no testing information was available for the student in grade 8 (which was the case for 34 percent of observations), the study team based the indicator on grade 7 information (which was available for 85 percent of missing cases). If grade 7 information was missing, the study team used grade 6 information (which was available for an additional 11 percent of missing cases). If grade 8, 7, and 6 information was unavailable, the student was considered not to have demonstrated proficiency.
    - For middle school science the constructed indicator measures whether a student scored proficient or above on the physical science assessment. This assessment is typically taken in grade 7, and all available data for science were for grade 7.
    - For high school science the constructed indicator measures whether a student scored proficient or above on the biology assessment. This assessment is typically taken in grade 10.
    - For high school math the constructed indicator measures whether a student scored proficient or above on either the algebra or the geometry assessment. These assessments are typically taken in grades 9 and 10, respectively.
    - No high school English language arts assessment was available for the relevant cohorts.

  Details on sensitivity analyses for alternate indicator constructions, as well as alignment with guidance in the Arkansas ESSA plan, are provided later in the appendix.

- **Discipline records.** These data include the number of suspensions and expulsions by grade. These tallies were used to create indicators for never suspended and never expelled during the middle school grades and during the high school grades.

- **High school course transcript records**. These data include information that was used to identify enrollment in an advanced course. Consistent with the Arkansas ESSA plan, Advanced Career Education, Advanced

Placement, and International Baccalaureate courses were classified as advanced. The constructed indicator measures whether a student enrolled in at least one advanced course. High school course transcript records were also used to construct an indicator measuring whether a student enrolled in at least one community service learning course. The Arkansas Department of Education includes two courses under this category: Community Service Learning and Leadership, and Service Learning. Finally, high school course transcript records were also used to identify students who earned an average grade point average (GPA) of 2.8 or higher, calculated as an average of yearly GPAs across the relevant high school years.

- **National Center for Education Statistics Common Core of Data**. Geographic locale records were used to identify each school district's locale as urban, suburban, town, or rural in the year in which a student was in grade 6 (U.S. Department of Education, n.d.).

Outcome variables. The three study postsecondary readiness and success outcome variables were constructed using data from Arkansas Department of Education, Arkansas Division of Higher Education, and National Student Clearinghouse.

- **Postsecondary readiness (ACT score of 19 or higher)**. Data on student scores on ACT college readiness exams are from the Arkansas Department of Education. The ACT is typically taken in grade 11, but some students took it multiple times, sometimes over multiple grades. The study team used the highest score available to identify whether students were college or career ready (the same practice followed by colleges), defined in the Arkansas ESSA plan benchmark as an ACT score of 19 or higher. Students with no or missing ACT scores were excluded from the descriptive statistics and estimation of models for postsecondary readiness. They were included in sensitivity analyses, with missing ACT scores coded as 0 (failing to attain postsecondary readiness).
- **Postsecondary success (college enrollment and persistence).** Data from the Arkansas Division of Higher Education include information on enrollment in higher education institutions in Arkansas, including public, private, nonprofit, and for-profit institutions. Enrollment was defined to include students pursuing or earning an associate's or a bachelor's degree or an academic or a technical certificate. National Student Clearinghouse (NSC) records include credential attainment records from institutions of higher education nationwide, including public, private, nonprofit, and for-profit. The study team used these data to generate an indicator of whether a student ever enrolled in college (enrollment) and another for whether the student was enrolled for more than one term or received a credential (persistence). Students who did not have records of college enrollment in more than one academic term were coded as not attaining persistence. Because NSC records include only credential attainment, all students in the NSC records were considered to have persisted according to the decision rules. Students who lacked records of enrollment in college in the data provided by the Arkansas Division of Higher Education and NSC were coded as 0 (failing to attain college enrollment or persistence).

For all three outcomes students who did not finish high school within eight years of beginning grade 6 were coded as failing to attain readiness and success, with the exceptions that students who were deceased; enrolled in home school, private school, or in another school out of state; or withdrew because of health problems were excluded from the analysis.

Data preparation

After duplicates were removed, the data file had 72,929 records. Next, 9,250 records were removed for students who were deceased; enrolled in home school, private school, or in another school out of state; or withdrew because of health problems. The final analytical sample included 63,679 students for each of the postsecondary

outcome models, and 37,930 students for the postsecondary readiness outcome model (only students with ACT records).[1]

To prepare the data to answer the three research questions, the study team performed four primary tasks:

1. Preparing data for merging, including making a single record per student ID for each data source (type of data).
2. Identifying students in grade 6 in 2008/09 and 2009/10 from the demographic data file and removing excluded cases.
3. Merging multiple data sources using student ID and year.
4. Aggregating variables and creating indicators for the analysis.

The stages are described in additional detail below.

Preparing data for merging. Student demographic, academic, attendance, and discipline records as received from the Arkansas Department of Education were sorted by grade level, retained when describing grades 6–12, and de-duplicated to a single annual record per student. Discipline variables were condensed by summing suspensions and expulsions separately across individual schools and types of discipline (in-school or out-of-school suspensions) by year. Attendance records were condensed first by removing pure duplicate records, then, for students who appeared at multiple schools in a year but at the same grade level, by using the student's days present and absent summed across all schools in a given year.

Demographic records were condensed by removing duplicate records and then by taking records only for the lowest grade level. Grade 6 records were used to establish gender, race/ethnicity, national school lunch program status, English learner status, and disability designation status unless there were multiple grade 6 records that conflicted with one another, in which case records for the next grade were used to reach a determination.

The GPA records had fewer than 0.1 percent duplicates, and so duplicates were dropped randomly until each student had a unique GPA record.

Identifying students in grade 6 in 2008/09 and 2009/10 from the demographics file and removing excluded cases. Students were identified as in grade 6 in 2008/09 (n = 36,477) or 2009/10 (n = 36,667) based on being listed as in grade 6 in the associated school year's demographic data file. When a student was listed as in grade 6 in both years' demographic files, only the 2008/09 record was kept (n = 215, total n = 72,929).

Students were dropped from the records if the coded reasons that they left school were "deceased," "enrolled in home school," "enrolled in private school," "enrolled in another school out of state," or "health problems," After this removal, the final n-size was 63,679.

Merging multiple data sources using student ID and year. All files were merged to the cohort level based on the anonymized student ID provided by the state. Because files were made unique by year and student ID before this step, these merges were all simple one-to-one merges without complication.

Aggregating variables and creating indicators for the analysis. Records were aggregated to the student or student/grade range level. Student demographic information was taken from the demographic data for grade 6, the time of cohort formation. Using later demographic information risks using an indicator observed after baseline—for example, exiting English learner status—as a predictor. As noted, the only exception was that the study team looked ahead to the next grade to adjudicate cases for which there were multiple grade 6 demographic records that conflicted with one another.

[1] Students with missing ACT scores were excluded from the models of postsecondary readiness, addressing research questions 2 and 3.

The study's predictive indicators were created following the Arkansas ESSA plan's listing and specifications for School Quality and Student Success indicators. In particular, the study included indicators for as many of the plan's Group A ("data for indicator available to calculate") and Group B ("data collection and calculation to be studied for future consideration") School Quality and Student Success elements as possible, given the administrative data available to the study team (Arkansas Department of Education, 2017).

To avoid extreme multicollinearity in the estimated logistic regression models, the study team decided to aggregate across the middle school years and across the high school years in constructing predictive indicators. Extreme multicollinearity (as would occur, for example, if English language arts or math proficiency levels from all available grades were entered into a model simultaneously as independent variables) yields estimated coefficients or marginal effects that are essentially uninterpretable. Multicollinearity considerations also led to the decision, after preliminary models were examined that included middle school and high school indicators together, not to present results for predictive models that included both sets of indicators simultaneously. A crosswalk of Arkansas ESSA plan School Quality and Student Success indicators by the current study's predictive indicators is displayed in table A1.

For models that included high school indicators as predictors of the postsecondary readiness outcome (ACT score of 19 or higher), the study team used information from grades 9 and 10 to construct high school indicators because the readiness outcome was assessed in the spring of grade 11 for most students. The study team used information from grades 9–12 to construct high school indicators to predict the postsecondary success outcomes because college enrollment and persistence were measured after high school completion.

Table A1. Crosswalk of Arkansas Every Student Succeeds Act (ESSA) plan School Quality and Student Success indicators (Groups A and B) by predictive indicators included in the study, 2008/09–2017/18

ESSA plan School Quality and Student Success group and indicator	Available and operationalized?			Notes on alignment with Arkansas ESSA plan
	Grades 6–8	Grades 9 and 10	Grades 9–12	
Group A				
Student absenteeism (chronic absenteeism)	Yes, aggregated across grades 6–8	Yes, aggregated across grades 9 and 10	Yes, aggregated across grades 9–12	One indicator for *X* <5 %; another indicator for 5% <= *X* <10%; baseline (or reference group) category is *X* = >10%, where *X* is the absence rate. Can also be expressed as percent of enrolled days present.
Science proficiency	Yes, ever proficient across grades 6–8	Yes, ever proficient across grades 9 and 10	Yes, ever proficient across grades 9–12	Middle school science test is a benchmark assessment typically given in grade 7; high school science test is a biology assessment typically given in grade 10; state-established proficiency thresholds translate to ESSA plan's language of "ready or exceeds on required state assessment."
Science growth	—	—	—	The study team does not have available value-added score metrics for study cohorts.

ESSA plan School Quality and Student Success group and indicator	Available and operationalized?			Notes on alignment with Arkansas ESSA plan
	Grades 6–8	Grades 9 and 10	Grades 9–12	
Reading at or above grade level	Yes, proficient in grade 8	—	—	Measured through state English language arts assessment; assessments administered in middle grades only for study cohorts. If there is no available information on grade 8 for a student, proficiency was assessed for grade 7; if there is no information on grade 7 for a student, proficiency was assessed for grade 6.
Meeting or exceeding state expectation of ACT composite score of 19	Yes, as college readiness outcome	Yes, as college readiness outcome	na	Measured as "ever attained 19 or higher during high school," consistent with ESSA plan; students who have missing ACT information are not included in this analysis.
Meeting or exceeding ACT readiness benchmark	—	—	—	Data were not available to the study team on these subject-specific scores. Also, they are the components for the college readiness outcome variable, so the study team would not want to model the subject-specific indicators in predicting a composite outcome.
Grade point average of 2.8 or better on 4.0 scale	—	Yes, aggregated across grades 9 and 10	Yes, aggregated across grades 9–12	Each high school grade level received a weight of 1 so that if a student repeated a grade, that grade does not have twice the weight of other grades in computing an average across grades; one of the most important variables in the prior literature (for example, Allensworth & Clark, 2020; Allensworth & Easton, 2005; Heppen & Therriault, 2008)
Earning credits in at least one community service learning course	—	na	Yes, aggregated across grades 9–12	Arkansas Department of Education classifies two courses under this category: Community Service Learning and Leadership and Service Learning. The study team employed this indicator only in predicting the postsecondary success outcomes because there is no clear theory why it should affect ACT performance.
On-time credits	—	—	—	Complete credit data were not provided, and thus these indicators could not be included in the analyses. The Arkansas ESAA plan indicates on-time status is represented by a student earning 5.5 credits by the end of grade 9, 11 credits by the end of grade 10, and 16.5 credits by the end of grade 11.
Computer science credits earned	—	—	—	Complete computer science coursetaking records were not provided, and thus these indicators could not be included in the analyses.

ESSA plan School Quality and Student Success group and indicator	Available and operationalized? Grades 6–8	Grades 9 and 10	Grades 9–12	Notes on alignment with Arkansas ESSA plan
Advanced Placement (AP), International Baccalaureate, Pre-AP, or Concurrent Credit (including Advanced Career Education) credits earned	na	na	Yes, aggregated across grades 9–12	Due to the timing of outcome measurement, the study team decided to model the readiness outcome (ACT score) on the basis of high school indicators reflecting grades 9 and 10. As most advanced coursetaking occurred in grades 11 and 12, the study team decided to use the advanced coursetaking indicator (based on grades 9–12) only in predicting success outcomes (college enrollment and persistence), not the readiness outcome. Enrollment in pertinent high school courses, rather than credits earned, was available for this study.
Group B				
Suspensions (both in school and out of school) and expulsions	Yes, suspensions aggregated across grades 6–8; and expulsions aggregated across grades 6–8	Yes, suspensions aggregated across grades 9 and 10; and expulsions aggregated across grades 9 and 10	Yes, suspensions aggregated across grades 9–12; and expulsions aggregated across grades 9–12	Arkansas ESSA plan describes school-level indicators based on reduction in rates of in-school and out-of-school suspensions and of expulsions; the study team operationalized suspensions and expulsions as consistently as possible with the Arkansas ESSA plan guidelines, using the available data.
Math proficiency	Yes, proficient in grade 8; where grade 8 achievement is missing, proficient in grade 7 is used; where grade 8 and 7 achievement are missing, achievement in grade 6 is used.	Yes, ever proficient on Algebra or Geometry assessment (at least one) across grades 9 and 10	Yes, ever proficient on Algebra or Geometry assessment (at least one) across grades 9–12	See notes, below, for row labeled "completion of above-grade-level coursetaking in math."
Foreign language in grade 8	—	na	—	Complete foreign language course-taking records were not provided, and thus these indicators could not be included in the analyses.
School closure of achievement gaps	—	—	—	School level indicators were not feasible for the present student-level study.
School reduction in disproportionate discipline rates for subgroups	—	—	—	School level indicators were not feasible for the present student-level study.

ESSA plan School Quality and Student Success group and indicator	Available and operationalized? Grades 6–8	Grades 9 and 10	Grades 9–12	Notes on alignment with Arkansas ESSA plan
Completion of above-grade-level coursetaking in math	—	—	—	Arkansas ESSA plan describes indicators based on students completing above-grade-level math courses and achieving "ready" or "above" on above-grade-level math assessment; the study team operationalized this indicator in a way that was most consistent with the Arkansas ESSA plan guidelines, using middle school assessments and high school results from algebra and geometry assessments available.
Career credential completion	—	—	—	Career credential completion data were not available for the current study.
Pre-apprenticeship or internship learning	—	—	—	Pre-apprenticeship or internship learning data were not available for the current study.
High school credits received in grades 5–9	—	—	—	Middle school credit data were not available for the current study.

na is not applicable. — is not available.
Source: Authors' analysis based on review of Arkansas Every Student Succeeds Act plan (Arkansas Department of Education, 2017).

Analysis methods

Research question 1. What percentage of Arkansas students attained the postsecondary readiness outcome (ACT score of 19 or higher) and success outcomes (college enrollment and persistence), and did attainment differ according to student background characteristics or status on postsecondary readiness indicators from middle school and high school? For this research question the study team examined the percentage of students who met or exceeded the Arkansas Department of Education benchmark score of 19 on the ACT, the percentage who enrolled in at least one term of college according to Arkansas Division of Higher Education records, and the percentage who enrolled in more than one term or who completed a credential according to Arkansas Division of Higher Education or NSC records.

To understand whether attainment differed by student background characteristics, the study team examined the percentage of students attaining each outcome by background characteristics. Univariate distributions for the student background characteristics are presented in table A2, and the percentages of students attaining each outcome by background characteristic are presented in table B1 in appendix B. To understand whether attainment differed by status on postsecondary readiness indicators, the study team examined the cross-tabulation of each of the three outcomes and each of the indicators. Univariate distributions for the middle school and high school indicators are presented in table A3, and the percentage of students attaining each outcome for these indicators is presented in tables B2–B5.

Table A2. Percentage of students with various background characteristics, by 2008/09 or 2009/10 grade 6 cohort and combined across cohorts

Background characteristic	2008/09 cohort	2009/10 cohort	Two cohorts combined
Student			
Male	51.3	50.9	51.1
Black	22.9	22.9	22.9
Hispanic	8.6	9.2	8.9
White	66.1	64.3	65.2
Other race/ethnicity	2.3	3.6	3.0
Eligible for national school lunch program	58.9	61.3	60.1
English learner student	5.7	6.2	5.9
With a designated disability	11.5	10.9	11.2
Entered grade 6 before age 13	80.8	82.4	81.6
School locale			
Urban	28.2	28.4	28.3
Suburban	8.8	9.1	9.0
Town	23.6	23.9	23.8
Rural	39.5	38.6	39.0

Note: The characteristics are based on student grade 6 information (2008/09 or 2009/10).
Source: Authors' analysis of data for 2008/09 and 2009/10 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table A3. Percentage of students who demonstrated each indicator, by 2008/09 or 2009/10 grade 6 cohort and combined cohorts, 2008/09–2017/18

Grade level and indicator	2008/09 cohort	2009/10 cohort	Two cohorts combined
Middle school, grades 6–8 (all outcomes)			
Proficient in English language arts	75.6	79.7	77.7
Proficient in math	65.2	70.2	67.7
Proficient in science	33.3	39.1	36.2
Present more than 95 percent of days enrolled	44.5	46.7	45.6
Present 91–95 percent of days enrolled	47.7	46.3	47
Never suspended	71.6	72.6	72.1
Never expelled	99.7	99.7	99.7
High school, grades 9 and 10 (readiness outcome)			
Proficient in math	68.2	67.7	68.0
Proficient in science	41.5	44.3	42.9
Grade point average of 2.8 or higher	45.8	47.1	46.5
Present more than 95 percent of days enrolled	57.8	57.2	57.5
Present 91–95 percent of days enrolled	32.7	33.0	32.8
Never suspended	75.9	75.6	75.7
Never expelled	99.5	99.7	99.6
High school, grades 9–12 (success outcome)			
Proficient in math	68.2	67.6	67.9
Proficient in science	41.5	44.3	42.9
Grade point average of 2.8 or higher	46.5	48.0	47.2
Enrolled in at least one advanced course	44.0	46.0	45.0
Enrolled in at least one community service learning course	3.7	3.9	3.8
Present 91–95 percent of days enrolled	51.8	52.4	52.1
Present more than 95 percent of days enrolled	37.6	36.8	37.2
Never suspended	66.1	65.3	65.7
Never expelled	99.3	99.4	99.3

Note: Information for the indicators of postsecondary readiness and success are pulled from yearly attendance, assessment, discipline, and transcript records. Readiness outcome = a score of 19 or higher on the ACT. For the success outcomes enrollment = enrolled for at least one term in a higher education institution, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6; persistence = enrolled in college for more than one term within eight years of beginning grade 6.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Research question 2. How accurately do postsecondary readiness indicators from middle school and high school predict attainment of the postsecondary readiness outcome (ACT score of 19 or higher) and success outcomes (college enrollment and persistence)? Does using the postsecondary readiness indicators improve the accuracy of these outcome predictors compared with using only student background characteristics? The data population was divided into training and testing samples for research question 2 to avoid selecting the most accurate model based on idiosyncrasies in the data that are not reflective of the broader population of Arkansas students. The study team used 70 percent of the cases to estimate optimal parameters of the models, which is commonly referred to

as model training (Hastie et al., 2009), and then tested the quality of the model on the remaining 30 percent of cases.

Next, regression and random forest models were used to predict postsecondary readiness and success using student characteristics and performance on the indicators and as a sensitivity check. For the logistic regression models the study team modeled findings as a linear function of the explanatory variables. In fitting the logistic regression models, the study team estimated coefficients of the linear function using the training dataset.

The random forest model used a collection of decision trees to predict a binary outcome. A majority voting rule was used to aggregate predictions from individual decision trees to the random forest prediction. The random forest model had two parameters: number of decision trees and number of randomly chosen variables to determine an optimal split at each decision node in the decision trees. Optimal values of both parameters of the random forest model were determined using the training dataset through cross-validation (see appendix C).

To assess model accuracy, a confusion matrix was used to measure model performance by comparing students' predicted successes and failures with actual successes and failures, to establish true positives, true negatives, false positives, and false negatives. True positives occur when a student who was predicted to attain readiness or success did attain readiness or success (table A4). True negatives occur when a student who was predicted not to attain readiness or success did not attain readiness or success. False positives occur when a student who was predicted to attain readiness or success did not attain readiness or success. False negatives occur when a student who was predicted not to attain readiness or success did attain readiness or success. Accuracy is the number of true positives plus the number of true negatives divided by the total sample size. All accuracy values are based on the testing sample. Appendix C describes supplementary analyses and findings that include estimates of postsecondary readiness and success using the random forest machine learning model.

Table A4. Confusion matrix for assessing model accuracy

	Predicted value	
Actual value	Readiness or success attained	Readiness or success not attained
Readiness or success attained	True positive	False negative
Readiness or success not attained	False positive	True negative

Note: True positives occur when a student who was predicted to attain readiness or success attained readiness or success. True negatives occur when a student who was predicted not to attain readiness or success did not attain readiness or success. False positives occur when a student who was predicted to attain readiness or success did not attain readiness or success. False negatives occur when a student who was predicted not to attain readiness or success attained readiness or success.
Source: Hastie et al. (2009).

Research question 3. After student background characteristics are controlled for, which middle school and high school indicators are the strongest predictors of the postsecondary readiness outcome (ACT score of 19 or higher) and success outcomes (college enrollment and persistence)? For research question 3 the study team used both logistic regression and random forest models.

The following logistic regression model was used:

$$LOGIT(y_{is}) = \mathbf{predictor}'_{is}\alpha + \boldsymbol{x}'_{\boldsymbol{i}}\beta + \varepsilon_{is}$$

where y_{is} represents the dummy variable associated with each of the outcomes for student *i* in school *s*; $\mathbf{predictor}_{is}$ represents a vector of the postsecondary readiness indicators of interest derived from the Arkansas ESSA plan (see table A1); $\mathbf{x}_i$ represents a vector of student background characteristics (including gender, race/ethnicity, district geographic locale, national school lunch program eligibility, English learner status, disability designation, age, and the urbanicity of a student's school); and ε_{is} is a random error term.

Postsecondary readiness indicators from middle school and high school were modeled separately as predictors to improve interpretability and reduce multicollinearity. First a model with only student background characteristics as predictors was estimated for a particular outcome. Then a model with student background characteristics and middle school indicators as predictors was estimated. Finally, a model with student background characteristics and high school indicators as predictors was estimated. Separately displaying associations of middle school indicators with outcomes and associations of high school indicators with outcomes is consistent with how educators and school systems will use the indicators to identify students who are off track or on track for attaining readiness and success. Standard errors are clustered at the district level. The study team also tested models using district-level fixed effects to control for persistent differences between districts (such as parents' education). These controls did not improve the accuracy of the models, so those results are not presented here.

For research questions 2 and 3 a sensitivity analysis was conducted for the postsecondary readiness outcome to examine how estimated models changed if students lacking ACT records were considered as not having attained postsecondary readiness and were included in model estimations. All the changes were insubstantial in that they did not change the interpretation of which variables are considered major.

Limitations

The study had several limitations. First, the associations reported for this study were based on cohorts of Arkansas students who entered grade 6 in 2008/09 and 2009/10. The historical nature of the data requires caution about interpretations or assumptions that similar relationships between indicators and outcomes exist for current K–12 students.

Second, the postsecondary readiness and success research points to a range of indicators that can be useful for predicting positive outcomes (Conley, 2012) beyond those used for this study. There is often substantial variation in postsecondary outcomes among students with similar academic performance in high school, suggesting that indicators in multiple domains are needed to more accurately predict postsecondary success (Beattie et al., 2018). Other indicators identified in the research include content knowledge (for example, knowledge in core subject areas and technical knowledge and skills); cognitive strategies (for example, problem formation, interpretation, and communication); learning skills and techniques (for example, soft skills such as ownership of learning, goal setting, persistence, time management, and self-monitoring); and knowledge and skills specific to the transition from high school (for example, understanding course sequences and career pathways, knowledge of financial aid and application options and procedures, and understanding college-level and workforce norms and expectations). Thus, while the current study examined the indicators that were available in Arkansas administrative data, additional indicators of readiness and success would likely be useful.

Third, in instances where the Arkansas ESSA plan did not give complete guidance, the study team had to make decisions on constructing indicators (see table A1). Different definitions or construction of indicators might lead to different results.

Fourth, the study team decided to exclude students from the analysis for five reasons other than graduation or dropping out (deceased; enrolled in home school, private school, or in another school out of state; or withdrew for health reasons). It is possible that these excluded students are different from the students who were analyzed in various ways; for example, student mobility in and of itself might represent a risk factor for being off track for postsecondary readiness. Therefore, the study findings should not be assumed to extend to students with the characteristics that led to exclusion from the analysis.

Fifth, the study team lacked data on students' workforce or military participation after high school, and thus the success outcomes focused solely on education outcomes after high school. That allowed for only a partial examination of what can be considered postsecondary success in the years following high school.

Finally, the findings are predictive but not causal. The findings from neither the logistic regression nor the random forest models should be considered to represent associations in which the indicators caused the outcomes with which they are associated. For example, students with higher middle school English language arts and math scores likely earn higher scores on the ACT college entrance exam not because any improvements in students' earlier test scores cause improvements in their later achievement but because underlying traits and skills (such as academic engagement and skills in goal setting) are enduring and correlated with both improved state achievement test scores and college entrance exam scores.

References

Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher, 49*(3), 198–211. https://doi.org/10.3102%2F0013189X20902110.

Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. University of Chicago Consortium on School Research. Retrieved February 1, 2021, from https://consortium.uchicago.edu/publications/track-indicator-predictor-high-school-graduation.

Arkansas Department of Education. (2017). *Every Student Succeeds Act: Arkansas State Plan*. http://dese.ade.arkansas.gov/public/userfiles/Public_School_Accountability/ESSA/Arkansas_ESSA_Plan_approved_01_16_2018_amended_03112019.pdf.

Beattie, G., Laliberté, J. W. P., & Oreopoulos, P. (2018). Thrivers and divers: Using non-academic measures to predict college success and failure. *Economics of Education Review*, *62*(1), 170–182.

Chang, H. N., Osher, D., Schanfield, M., Sundius, J., & Bauer, L. (2019, September). *Using chronic absence data to improve conditions for learning*. Attendance Works & American Institutes for Research. https://eric.ed.gov/?id=ED602448.

Conley, D. T. (2012). *A complete definition of college and career readiness.* Educational Policy Improvement Center. https://eric.ed.gov/?id=ED537876.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Heppen, J. B., & Therriault, S. B. (2008). *Developing early warning systems to identify potential high school dropouts* (Issue Brief). National High School Center. https://eric.ed.gov/?id=ED521558.

Jordan, P. W., & Miller, R. (2017). *Who's in: Chronic absenteeism under the Every Student Succeeds Act*. FutureED, Georgetown University. https://www.future-ed.org/wp-content/uploads/2017/09/REPORT_Chronic_Absenteeism_final_v5.pdf.

U.S. Department of Education, National Center for Education Statistics. (n.d.). Elementary/secondary information system. Retrieved February 1, 2021, from https://nces.ed.gov/ccd/elsi/.

Appendix B. Supporting tables

This appendix includes supplementary tables that support the findings in the main report.

Table B1. Percentage of students from the 2008/09 and 2009/10 grade 6 cohorts who attained postsecondary readiness and success outcomes within eight years of beginning grade 6, by student background characteristics, 2007/08–2017/18

Background characteristic	Readiness outcome[a]	Success outcome	
	ACT score of 19 or higher	College enrollment[b]	College persistence[c]
Overall	**64.5**	**58.0**	**49.1**
Gender			
Male	64.2	52.3	43.1
Female	64.8	64.1	55.5
Race/ethnicity			
Black	33.3	51.7	41.4
Hispanic	50.6	43.1	37.3
White	74.6	62.4	53.5
Other	74.2	55.7	48.1
Eligible for the national school lunch program			
Yes	49.2	47.9	38.7
No	79.3	73.3	64.8
English learner student			
Yes	38.4	35.1	30.2
No	65.7	59.5	50.3
Has a disability designation			
Yes	22.9	30.2	22.5
No	66.7	61.6	52.5
Entered grade 6 before age 13			
Yes	67.5	63.1	53.9
No	41.3	35.7	28.1
District locale			
Urban	59.8	47.6	38.9
Suburban	70.8	60.6	52.1
Town	64.3	62.8	53.4
Rural	64.3	61.7	52.7

a. Analytic sample for the readiness outcome excludes students who did not take the ACT. The demographic characteristics are as of grade 6.
b. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.
c. Enrolled in college for more than one term within eight years of beginning grade 6.
Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B2. Percentage of students from the 2008/09 and 2009/10 grade 6 cohorts who attained the postsecondary readiness outcome within eight years of beginning grade 6, by middle school indicator, 2007/08–2017/18

Middle school indicator	Readiness outcome[a]: ACT score of 19 or higher; percent	Difference from reference group (percentage point)
Overall	**64.5**	
Proficient in English language arts		
Yes	71.1	61.4
No[b]	9.7	
Proficient in math		
Yes	76.9	62.5
No[b]	14.4	
Proficient in science		
Yes	91.3	50.2
No[b]	41.1	
Attendance category		
Present more than 95 percent of days enrolled	66.5	11.8
Present 91–95 percent of days enrolled	62.9	8.2
Present 90 percent or fewer of days enrolled (chronic absenteeism)[a]	54.7	
Ever suspended		
Yes	42.6	–27.1
No[b]	69.7	
Ever expelled		
Yes	18.5	–46.1
No[b]	64.6	

Note: The middle school grades are 6–8 for the definition and construction of the indicators for the readiness outcome.
a. Analytic sample for the readiness outcome excludes students who did not take the ACT.
b. Reference group for calculating difference.
Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B3. Percentage of students from the 2008/09 and 2009/10 grade 6 cohorts who attained the postsecondary success outcomes (college enrollment and persistence) within eight years of beginning grade 6, by middle school indicator, 2007/08–2017/18

	Success outcome			
Middle school indicator	College enrollment[a] (percent)	Difference from reference group (percentage points)	College persistence[b] (percent)	Difference from reference group (percentage points)
Overall	**58.0**		**49.1**	
Proficient in English language art				
Yes	66.3	36.9	57.1	35.7
No[c]	29.4		21.4	
Proficient in math				
Yes	68.4	31.9	59.5	32.0
No[c]	36.5		27.5	
Proficient in science				
Yes	74.7	26.1	65.8	26.1
No[c]	48.6		39.7	
Attendance category				
Present more than 95 percent of days enrolled	67.0	37.0	58.2	35.3
Present 91–95 percent of days enrolled	53.8	23.8	44.4	21.5
Present 90 percent or fewer of days enrolled (chronic absenteeism)[c]	30.0		22.9	
Ever suspended				
Yes	40.1	–24.9	30.8	–25.4
No[c]	65.0		56.2	
Ever expelled				
Yes	10.6	–47.6	7.1	–42.2
No[c]	58.2		49.3	

Note: The middle school grades are 6–8 for the definition and construction of indicators for the success outcomes.
a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.
b. Enrolled in college for more than one term within eight years of beginning grade 6.
c. Reference group for calculating difference.
Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B4. Percentage of students from the 2008/09 and 2009/10 grade 6 cohorts who attained the postsecondary readiness outcome, by high school indicator, 2007/08–2017/18

High school indicator	Readiness outcome[a] ACT score of 19 or higher; percent	Difference from reference group (percentage points)
Overall	**64.5**	
Proficient in math		
Yes	74.2	47.5
No[b]	26.7	
Proficient in science		
Yes	89.4	56.2
No[b]	33.2	
Grade point average of 2.8 or higher		
Yes	81.1	46.9
No[b]	34.2	
Attendance category		
Present more than 95 percent of days enrolled	67.9	18.8
Present 91–95 percent of days enrolled	59.0	9.9
Present 90 percent or fewer of days enrolled (chronic absenteeism)[b]	49.1	
Ever suspended		
Yes	38.4	–30.9
No[b]	69.3	
Ever expelled		
Yes	30.3	–34.3
No[b]	64.6	

Note: The high school grades are 9–10 for the for the definition and construction of the indicators for the readiness outcome.
a. Analytic sample for readiness outcome excludes students who did not take the ACT.
b. Reference group for calculating difference.
Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B5. Percentage of students from the 2008/09 and 2009/10 grade 6 cohorts who attained the postsecondary success outcomes (college enrollment and persistence) within eight years of beginning grade 6, by high school indicator, 2007/08–2017/18

High school indicator	Success outcomes			
	College enrollment[a] (percent)	Difference from reference group (percentage points)	College persistence[b] (percent)	Difference from reference group (percentage points)
Overall	**58.0**		**49.1**	
Proficient in math				
Yes	68.2	31.7	59.0	30.8
No[c]	36.5		28.2	
Proficient in science				
Yes	75.6	30.7	66.6	30.6
No[c]	44.9		36.0	
Grade point average of 2.8 or higher				
Yes	80.6	42.7	72.4	44.0
No[c]	37.9		28.4	
Enrolled in at least one advanced course				
Yes	83.2	45.8	74.5	46.1
No[c]	37.4		28.4	
Enrolled in at least on community service learning course				
Yes	64.9	7.1	56.6	7.8
No[c]	57.8		48.8	
Attendance category				
Present more than 95 percent of days enrolled	67.5	40.6	59.8	39.4
Present 91–95 percent of days enrolled	57.7	30.8	47.4	27.0
Present 90 percent or fewer of days enrolled (chronic absenteeism)[c]	26.9		20.4	
Ever suspended				
Yes	42.5	–23.7	32.4	–25.5
No[c]	66.2		57.9	
Ever expelled				
Yes	16.8	–41.5	10.8	–38.6
No[c]	58.3		49.4	

Note: The high school grades are 9–12 for the definition and construction of indicators for the success outcomes.
a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.
b. Enrolled in college for more than one term within eight years of beginning grade 6.
c. Reference group for calculating difference.
Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B6. Accuracy rate and true positive, false positive, true negative, and false negative rates for postsecondary readiness and success outcomes for the logistic regression models estimated for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Readiness outcome	Success outcome	
Logistic regression model	ACT score of 19 or higher	College enrollment[a]	College persistence[b]
With background characteristics only			
Accuracy rate	**73.7**	**67.3**	**65.8**
True positive rate	86.7	82.4	66.7
True negative rate	50.2	46.4	64.9
False positive rate	49.8	53.6	35.1
False negative rate	13.3	17.6	33.3
With middle school indicators and background characteristics			
Accuracy rate	**82.1**	**71.7**	**69.9**
True positive rate	89.0	82.3	72.5
True negative rate	69.6	57.1	67.4
False positive rate	30.4	42.9	32.6
False negative rate	11.0	17.7	27.5
With high school indicators and background characteristics			
Accuracy rate	**82.9**	**75.8**	**75.0**
True positive rate	88.0	79.1	73.4
True negative rate	73.8	71.2	76.6
False positive rate	26.2	28.8	23.4
False negative rate	12.0	20.9	26.6

Note: True positives occur when a student who was predicted to attain readiness or success attained readiness or success (see table A4 in appendix A). True negatives occur when a student who was predicted not to attain readiness or success did not attain readiness or success. False positives occur when a student who was predicted to attain readiness or success did not attain readiness or success. False negatives occur when a student who was predicted not to attain readiness or success attained readiness or success. Accuracy is the number of true negatives plus the number of true positives divided by the total sample size. The middle school indicators are associated with grades 6–8 for all three outcomes. The high school indicators are associated with grades 9 and 10 for the readiness outcome and 9–12 for the success outcomes.

a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.

b. Enrolled in college for more than one term within eight years of beginning grade 6.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B7. Marginal effects estimates of logistic regression coefficients for middle school indicators and the postsecondary readiness outcome (ACT score of 19 or higher) modeled for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Readiness outcome
Indicator	ACT score of 19 or higher
Proficient in English language arts	23.8*** (1.13)
Proficient in math	27.6*** (0.78)
Proficient in science	27.2*** (0.52)
Present 91–95 percent of days enrolled	–1.8 (0.95)
Present more than 95 percent of days enrolled	–0.8 (1.04)
Never suspended	5.1*** (0.57)
Never expelled	17.4 (10.1)

*** Significant at $p < .001$.

Note: Results are based on marginalization of logistic regression models generated using the mfx logistic regression package in R and the option that calculates the partial effect for each observation unit and then averages them. All models include the full set of control variables, including race/ethnicity, gender, national school lunch program eligibility, English learner status, disability designation, whether the student entered grade 6 before age 13, and district locale.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B8. Marginal effects estimates of logistic regression coefficients for middle school indicators and the postsecondary success outcomes (college enrollment and persistence) modeled for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Success outcome	
Indicator	College enrollment[a]	College persistence[b]
Proficient in English language arts	12.4*** (0.77)	12.6*** (0.71)
Proficient in math	10.2*** (0.59)	11.6*** (0.59)
Proficient in science	9.3*** (0.62)	8.4*** (0.66)
Present 91–95 percent of days enrolled	12.3*** (0.82)	12.4*** (0.86)
Present more than 95 percent of days enrolled	19.2*** (0.98)	19.7*** (1.03)
Never suspended	10.7*** (0.71)	11.4*** (0.68)
Never expelled	28.6*** (4.37)	27.3*** (3.66)

*** Significant at $p < .001$.

Note: Results are based on marginalization of logistic regression models generated using the mfx logistic regression package in R and the option that calculates the partial effect for each observation unit and then averages them. All models include the full set of control variables, including race/ethnicity, gender, national school lunch program eligibility, English learner status, disability designation, whether the student entered grade 6 before age 13, and district locale.

a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.

b. Enrolled in college for more than one term within eight years of beginning grade 6.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B9. Marginal effects estimates of logistic regression coefficients for high school indicators and the postsecondary readiness outcome (ACT score of 19 or higher) modeled for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Readiness outcome
Indicator	ACT score of 19 or higher
Proficient in math	14.4***
	(1.20)
Proficient in science	32.5***
	(0.67)
Grade point average of 2.8 or higher	16.6***
	(0.71)
Present 91–95 percent of days enrolled	0.2
	(0.78)
Present more than 95 percent of days enrolled	1.7*
	(0.82)
Never suspended	2.8***
	(0.72)
Never expelled	3.5
	(6.95)

* Significant at $p < .05$; *** significant at $p < .001$.

Note: Results are based on marginalization of logistic regression models generated using the mfx logistic regression package in R and the option that calculates the partial effect for each observation unit and then averages them. All models include the full set of control variables, including race/ethnicity, gender, national school lunch program eligibility, English learner status, disability designation, whether the student entered grade 6 before age 13, and district locale. The high school indicators are associated with grades 9 and 10.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table B10. Marginal effects estimates of logistic regression coefficients for high school indicators and postsecondary success outcomes (college enrollment and persistence) modeled for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Success outcome	
Indicator	College enrollment[a]	College persistence[b]
Proficient in math	9.0*** (0.57)	8.4*** (0.60)
Proficient in science	2.9*** (0.55)	2.0*** (0.56)
Grade point average of 2.8 or higher	15.6*** (0.94)	16.8*** (0.90)
Enrolled in at least one advanced course	24.1*** (0.90)	23.0*** (0.86)
Enrolled in at least one community service learning course	1.1 (1.77)	1.0 (1.51)
Present 91–95 percent of days enrolled	11.4*** (0.73)	10.8*** (0.77)
Present more than 95 percent of days enrolled	14.6*** (0.94)	15.0*** (0.99)
Never suspended	2.4*** (0.53)	4.2*** (0.61)
Never expelled	16.5*** (2.50)	17.0*** (2.60)

*** Significant at $p < .001$.

Note: Results are based on marginalization of logistic regression models generated using the mfx logistic regression package in R and the option that calculates the partial effect for each observation unit and then averages them. All models include the full set of control variables, including race/ethnicity, gender, national school lunch program eligibility, English learner status, disability designation, whether the student entered grade 6 before age 13, and district locale. The high school indicators are associated with grades 9–12.

a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.

b. Enrolled in college for more than one term within eight years of beginning grade 6.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Reference

U.S. Department of Education, National Center for Education Statistics. (n.d.). Elementary/secondary information system. Retrieved February 1, 2021, from https://nces.ed.gov/ccd/elsi/.

Appendix C. Alternative model results

Previous research has shown that machine learning generally, and random forest models in particular, can achieve higher accuracy than logistic regression models in predicting certain education outcomes, such as high school dropout (Knowles, 2015). This study used both logistic regression and random forest models to estimate results, to ensure that they are as accurate as possible, given the data available.

Random forest model accuracy was consistent with logistic regression findings

As with the logistic regression model results, the overall accuracy of the random forest models was moderately improved when middle school and high school indicators were included in the model along with student background characteristics (table C1). For all three postsecondary outcomes the accuracy of the random forest models was within 2 percentage points of the accuracy of the logistic regression models (see table B6 in appendix B).

Table C1. Accuracy rate and true positive, false positive, true negative, and false negative rates for postsecondary readiness and success outcomes for the random forest models estimated for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Readiness outcome	Success outcome	
	ACT score of 19 or higher	College enrollment[a]	College persistence[b]
With background characteristics			
Accuracy rate	**73.6**	**67.8**	**66.1**
True positive rate	87.1	80.8	69.5
True negative rate	49.3	49.7	62.8
False positive rate	50.7	50.3	37.2
False negative rate	12.9	19.2	30.5
With middle school indicators and background characteristics			
Accuracy rate	**81.7**	**71.6**	**69.8**
True positive rate	88.6	83.0	72.5
True negative rate	69.2	55.7	67.2
False positive rate	30.8	44.3	32.8
False negative rate	11.4	17.0	27.5
With high school indicators and background characteristics			
Accuracy rate	**82.0**	**75.3**	**73.4**
True positive rate	87.4	79.4	72.4
True negative rate	72.4	69.7	74.5
False positive rate	27.6	30.3	25.5
False negative rate	12.6	20.6	27.6

Note: True positives occur when a student who was predicted to attain readiness or success attained readiness or success (see table A4 in appendix A). True negatives occur when a student who was predicted not to attain readiness or success did not attain readiness or success. False positives occur when a student who was predicted to attain readiness or success did not attain readiness or success. False negatives occur when a student who was predicted not to attain readiness or success attained readiness or success. Accuracy is the number of true negatives plus the number of true positives divided by the total sample size. The middle school indicators are associated with grades 6–8 for all three outcomes. The high school indicators are associated with grades 9 and 10 for the readiness outcome and 9–12 for the success outcomes.

a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.

b. Enrolled in college for more than one term within eight years of beginning grade 6.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Random forest model marginal effects estimates were mostly consistent with logistic regression estimates

For a given indicator and postsecondary outcome, the fitted random forest model was used to compute the predicted probability of the outcome when the indicator was set to 0 and when the indicator was set to 1 for each student. The average difference between the two across all students corresponds to the marginal effect of the indicator on the postsecondary outcome. This characterization of marginal effects is identical to that of marginal effects computed using logistic regression models.

For the middle school indicators the study found marginal effects of at least 10 percentage points for at least one postsecondary readiness or success outcome for proficiency in English language arts, math, and science; never suspended; and present more than 95 percent of days enrolled (table C2). The marginal effects estimates for these indicators suggest that the increased probability of attaining the readiness and success outcomes for a student who scored proficient or above in grade 8 was 30–33 percentage points for proficiency in English language arts, 20–44 percentage points for proficiency in math, 9–22 percentage points for proficiency in science, and 7–20 percentage points for never being suspended. Students who were never expelled had a 5–15 percentage point increased probability of attaining the readiness and success outcomes. Students attending school more than 95 percent of days enrolled had a 0–16 percentage point increased probability of attaining the readiness and success outcomes.

The marginal effects of the middle school indicators were smaller for the logistic regression models than for the random forest models. For example, the logistic regression models estimated that a student who achieved math proficiency in grade 8 had as large as a 28 percentage point increased probability of attaining the readiness and success outcomes, whereas the random forest model estimated an increased probability as large as 44 percentage points. However, the order of importance of the variables based on effect size was mostly similar across both models.

For the high school indicators the study found marginal effects of at least 10 percentage points for at least one postsecondary readiness or success outcome for proficiency in math and science, grade point average (GPA), enrollment in at least one advanced course, and present for more than 95 percent of days enrolled (table C3). The marginal effects estimates for these indicators suggest that the increased probability of attaining the readiness and success outcomes for a student who scored proficient or above in high school was 12–20 percentage points for proficiency in math and 8–33 percentage points for proficiency in science. The increased probability of attaining the readiness and success outcomes was 20–26 percentage points for student who earned a GPA of 2.8 or higher, 30–31 percentage points for a student who enrolled in at least one advanced course, and 2-16 percentage point for a student who was present more than 95 percent of days enrolled.

The sizes of the marginal effects of the high school indicators were similar to those of the logistic regression models, except for the never expelled indicator. The logistic regression models estimated that a student who was never expelled during high school had as large as a 17 percentage point increased probability of attaining the readiness and success outcomes, whereas the random forest model estimated a smaller increased probability of 6–7 percentage points.

These differences between the two types of models could have arisen because the random forest model accommodates interactions, or multiplicative associations, among two, three, or more indicators, whereas the logistic regression model did not include any interaction terms. When interaction terms are not specified in a predictive model, the association with a specific (that is, first-order) predictor can be inflated as a result of model specification.

Table C2. Middle school indicators: Marginal effects estimates of random forest model for middle school indicators and the postsecondary readiness and success (enrollment and persistence) outcomes modeled for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Readiness outcome	Success outcomes	
Indicator	ACT score of 19 or higher	College enrollment[a]	College persistence[b]
Proficient in English language arts	33.0	30.0	30.0
Proficient in math	43.7	20.2	27.1
Proficient in science	21.8	8.7	12.5
Present more than 95 percent of days enrolled	0.3	10.6	15.8
Present 91–95 percent of days enrolled	–0.5	1.0	4.2
Never suspended	6.9	12.9	20.2
Never expelled	8.1	14.8	4.9

Note: Results are based on marginalization of random forest models, which calculates the partial effect for each observation unit and then averages them. Unlike the marginal effects estimates in the logistic regression models, it is not straightforward to compute standard errors, and thus *p*-values, when using random forest models. All models include the full set of control variables, including race/ethnicity, gender, national school lunch program eligibility, English learner status, disability designation, whether the student entered grade 6 before age 13, and district locale. Present at least 91 percent of days is accompanied by an indicator for whether days absent was missing.

a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.

b. Enrolled in college for more than one term within eight years of beginning grade 6.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

Table C3. Marginal effects estimates of random forest model for high school indicators and postsecondary readiness and success (enrollment and persistence) outcomes modeled for students from the 2008/09 and 2009/10 grade 6 cohorts, 2007/08–2017/18

	Readiness outcome	Success outcome	
Indicator	ACT score of 19 or higher	College enrollment[a]	College persistence[b]
Proficient in math	19.9	11.9	11.6
Proficient in science	33.3	8.3	7.5
Grade point average of 2.8 or higher	19.5	23.8	26.4
Enrolled in at least one advanced course	na	30.4	30.6
Enrolled in at least one community service learning course	na	1.0	0.5
Present more than 95 percent of days enrolled	2.3	9.1	15.8
Present 91–95 percent of days enrolled	0.2	3.4	8.6
Never suspended	4.7	5.5	7.5
Never expelled	8.4	7.6	5.6

na is not applicable because the variable was not included in the model predicting the readiness outcome.

Note: Results are based on marginalization of random forest models, which calculates the partial effect for each observation unit and then averages them. Unlike the marginal effects estimates in the logistic regression models, it is not straightforward to compute standard errors, and thus *p*-values, when using random forest models. All models include the full set of control variables, including race/ethnicity, gender, national school lunch program eligibility, English learner status, disability designation, whether the student entered grade 6 before age 13, and district locale. Regressions also control for the full set of middle school indicators shown in table C2, including proficiency in English language arts, science, and math, present at least 91 percent of days enrolled in middle school; and indicators for suspension and expulsion.

a. Enrolled in college for at least one term, regardless of the degree or certificate being pursued or attained, within eight years of beginning grade 6.

b. Enrolled in college for more than one term within eight years of beginning grade 6.

Source: Authors' analysis of data for 2008/09–2017/18 from the Arkansas Department of Education, Arkansas Division of Higher Education, National Student Clearinghouse, and National Center for Education Statistics Common Core of Data (U.S. Department of Education, n.d.).

References

Knowles, J. E. (2015). Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. *Journal of Educational Data Mining, 7*(3), 18–67. https://eric.ed.gov/?id=EJ1115326.

U.S. Department of Education, National Center for Education Statistics. (n.d.). Elementary/secondary information system. Retrieved February 1, 2021, from https://nces.ed.gov/ccd/elsi/.