Making Connections

July 2016

Identifying early warning indicators in three Ohio school districts

David Stuit
Basis Policy Research

Mindee O'Cummings
Heather Norbury
Jessica Heppen
Sonica Dhillon
Jim Lindsay
Bo Zhu
American Institutes for Research

In collaboration with the Midwest Dropout Prevention Research Alliance

Summary

In partnership with the Midwest Dropout Prevention Research Alliance the study team used student-level data and a five-step process to identify the most accurate indicators of students' failure to graduate from high school on time. Student-level data came from attendance records, transcripts, and discipline records of grade 8 and 9 students in three Ohio school districts. The study found that the most accurate early warning indicators of students being off track for graduating on time vary by school district and grade level. Overall, the most accurate indicators in both grades were based on coursework (grade point average and number of credits earned). On average, indicators were more accurate in grade 9 than in grade 8. Other districts may be able to use the methods described in this report to identify early warning indicators for their grade 8 and 9 students.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL MIDWEST
Regional Educational Laboratory
At American Institutes for Research

U.S. Department of Education
John B. King, Jr., *Secretary*

Institute of Education Sciences
Ruth Neild, *Deputy Director for Policy and Research*
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Joy Lesnick, *Acting Commissioner*
Amy Johnson, *Action Editor*
Elizabeth Eisner, *Project Officer*

REL 2016–118

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

July 2016

This report was prepared for the Institute of Education Sciences (IES) under ED-IES-12-C-0004 by Regional Educational Laboratory Midwest administered by American Institutes for Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Stuit, D., O'Cummings, M., Norbury, H., Heppen, J., Dhillon, S., Lindsay, J., & Zhu, B. (2016). *Identifying early warning indicators in three Ohio school districts* (REL 2016–118). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

A growing number of school districts are using early warning systems in their strategy for improving rates of student on-time graduation. Such systems use academic and behavioral indicators from student-level data to identify students who are at risk of not graduating high school on time. Once students are identified, the school district can provide them with supplemental supports (for example, supplemental instruction or counseling) to get them back on track to graduate on time.

The Midwest Dropout Prevention Research Alliance is composed of representatives of state education agencies, intermediate education agencies, and school districts in Regional Educational Laboratory (REL) Midwest Region states who share a commitment to reduce high school dropout rates. Alliance members wanted to know whether school districts should develop their own early warning systems or adopt those that have been validated in other settings (for example, the system developed by the Consortium for Chicago School Research or the system developed by the National High School Center). The alliance partnered with REL Midwest to address this question.

The study team followed a five-step process and used student data for two cohorts of grade 8 and 9 students in three Ohio school districts (referred to as Districts A, B, and C) to identify indicators that predict failure to graduate on time. The three districts varied in size, demographic composition, and locale. Two districts serve large cities with a population greater than 250,000, while the third district serves a town near an urban area. One of the urban districts has more than 40,000 students, while the other districts each have 5,000–10,000 students. The percentage of students qualifying for the federal school lunch program (a proxy for low income) also varied, ranging from about 40 percent to more than 90 percent. The four-year graduation rate for the three districts ranged from 56 percent to 91 percent.

Student-level data on attendance, achievement, coursework, and discipline were used to construct a set of indicators for grade 8 and 9 students that were candidates for inclusion in each district's early warning system. The number of indicators available from each district varied. The following indicators were included in the analysis for at least one district: end-of-year attendance rate, grade point average, number of credits earned, number of failing grades, number of failing grades in core courses, number of suspensions, and reading and math scores on the Ohio Achievement Assessment (grade 8 only).

Students were designated as either on track or off track based on whether their performance on each candidate indicator fell above or below the optimal cutpoint for predicting whether they would graduate on time. The study team analyzed how the optimal cutpoints on the candidate indicators varied across districts and grades. After applying the optimal cutpoints to the candidate indicators, the study team conducted a series of statistical tests to eliminate candidates that were not consistently predictive of failure to graduate on time when applied to 100 simulated cohorts of grade 8 and grade 9 students. The study team then identified the indicators with the highest correct prediction rates, the lowest false alarm rates, and best overall accuracy (best balance between correct predictions and false alarms) among indicators that passed the consistency tests. Finally, the study team looked at the degree to which the accuracy of the indicators varied across districts and grades.

The analyses were restricted to grade 8 and grade 9 data for students who were first-time freshmen in the districts in 2006/07 or 2007/08 and excluded students who entered the districts after grade 9. Students in the 2006/07 cohort graduated in 2010, and students in the 2007/08 cohort graduated in 2011.

Certain indicators were more accurate predictors of failure to graduate on time in some districts than other indicators were, and the optimal cutpoints for classifying students as on track or off track for graduation differed across districts. Of student data for grades 8 and 9 the end-of-year attendance rate was the only consistent predictor of failure to graduate on time in all three districts. Reading scores from the Ohio Achievement Assessment were consistent predictors for grade 8 students in all three districts. The most accurate indicators in both grade 8 and grade 9 were based on coursework (grade point average and number of credits earned). Consistent with prior research, failing more than one class and being suspended one or more times were also strong predictors of failure to graduate on time. On average, indicators were more accurate in grade 9 than in grade 8.

Given the variability across school districts and grade levels in optimal cutpoints, in consistency of predicting failure to graduate on time, and in relative accuracy of indicators to predict failure to graduate on time, the findings suggest that it is important for school districts to examine and analyze their own student-level data in order to develop their own early warning system. The methods used in this study can help districts identify the best off-track indicators and indicator cutpoints for their early warning system.

Contents

Why this study?

Information from early warning systems can help educators target resources and interventions to students at the greatest risk of not graduating or not graduating on time. But how can districts determine which types of student data to use for their early warning indicators? This study identified valid grade 8 and 9 early warning indicators developed from datasets collected from three school districts that vary in size, urbanicity, and the characteristics of their student populations. The findings include information on the accuracy of each district's indicators for predicting whether students will fail to graduate within four years. Evidence from this study may help these districts identify an accurate set of early warning indicators, and the indicator identification process described here can inform the efforts of state, district, and school leaders who wish to develop their own early warning systems as a means of keeping students on track to graduate.

Evidence from this study may help districts identify an accurate set of early warning indicators, and the indicator identification process described here can inform the efforts of state, district, and school leaders who wish to develop their own early warning systems as a means of keeping students on track to graduate

Early warning systems help identify students who are at risk of not graduating on time and need extra support

Early warning systems use data systematically to identify students who are at risk of not graduating on time. Students identified early can be matched with interventions to help them return to the on-time graduation track (Heppen & Therriault, 2008; Jerald, 2006; Kennelly & Monrad, 2007; Neild, Balfanz, & Herzog, 2007; Pinkus, 2008). The push for early warning systems is motivated by research on academic and behavioral predictors of students dropping out of school (Allensworth & Easton, 2005, 2007; Balfanz, Herzog, & Mac Iver, 2007; Neild & Balfanz, 2006; Silver, Saunders, & Zarate, 2008). Although the evidence base for early warning systems is still developing, experts on dropout prevention consider these systems a promising approach—or a necessary prerequisite—to effective dropout prevention (Dynarski et al., 2008). See appendix A for a review of relevant literature on early warning systems.

Researchers have identified a set of core early warning indicators

Analysis of data from large urban districts has enabled researchers to identify indicators that predict whether middle school and grade 9 students will graduate from high school on time. For middle school students, attendance, course grades, and behavior such as out-of-school suspensions have been frequently identified as early indicators of high school graduation outcomes (Balfanz & Herzog, 2005; Neild & Balfanz, 2006; Balfanz, 2009). For high school students, attendance, course performance, credit attainment, and, in some cases, state assessment scores, grade retention, and behavior have been frequently identified as early indicators (Allensworth & Easton, 2005, 2007; Neild & Balfanz, 2006; Roderick, 1993; Silver et al., 2008).

District personnel must be able to calculate indicators easily and communicate them to educators and parents. The task of communicating whether students are on track or off track to graduate on time is made easier by converting continuous indicators—that is data elements that can have a range of possible numeric values, from low to high, such as a grade point average or test score—into binary indicators that classify students as either on track or off track based on whether their score falls above or below a particular cutpoint. The location of the cutpoint on the continuous scale of the indicator is set at the value that most accurately distinguishes between students who are at risk of not graduating on time and those who are not at risk (see definitions of terms in box 1). For example,

Box 1. Definitions of key terms

Candidate indicator. A data element representing grade 8 or 9 student academic performance or behavior that may predict not graduating on time. Candidate indicators in the study included count data (number of credits earned, number of failed courses, number of suspensions), scaled data (test scores, grade point averages), and proportions (attendance rates). Each candidate indicator was converted from its original scale to a binary (yes/no; high/low) scale by identifying the optimal cutpoint.

Correct off-track prediction rate. The proportion of nongraduates whom the grade 8 or 9 binary indicator correctly identified (flagged) as being at risk of not graduating on time. For example, if the binary indicator for attendance flagged all students with an attendance rate less than 90 percent and 75 percent of nongraduates had an attendance rate below this optimal cutpoint, the correct off-track prediction rate for attendance would be 75 percent.

Failure to graduate on time. The outcome measure used in this study. Students in each grade 8 and 9 cohort are classified as not graduating on time if they did not receive a high school diploma within four years of beginning high school. Students who do not graduate within four years are classified as nongraduates even though they may later graduate.

False alarm rate. The percentage of students who graduate on time whom the grade 8 or 9 binary indicator incorrectly flagged. For example, if 35 percent of students who graduate on time had an attendance rate below the indicator cutpoint of 90 percent, the false alarm rate would be 35 percent.

Optimal cutpoint. A specific value on the original scale of a grade 8 or 9 indicator that separates students who are at risk of not graduating on time from those who are not at risk. For example, if 90 percent is the optimal cutpoint for the grade 9 attendance indicator, students with an attendance rate at or below 90 percent could be classified as at risk of failure to graduate on time. For each indicator, the optimal cutpoint is the value that produces the highest rate of correct off-track prediction and lowest rate of mistaken predictions (that is, false alarms).

Overall accuracy. A statistic that measures the balance between an indicator's correct off-track prediction rate and false alarm rate on a scale from .50 to 1.00, with a higher value indicating greater accuracy (more correct off-track predictions and fewer false alarms). A score of .50 means that the indicator is no better at predicting which students will not graduate than random guessing. A score of 1.00 means that the indicator perfectly predicts failure to graduate on time (100 percent correct off-track prediction rate and 0 percent false alarm rate). The formal name for this statistic is the area under the curve; it is based on a statistical technique called receiver operating characteristic curve analysis (see appendix B for details).

converting grade 9 students' grade point average into a binary indicator (on track or off track) involves locating the particular grade point average that best separates students who graduate on time from students who do not. If that cutpoint is 2.0, then students with a grade point average lower than 2.0 are classified as off track for on-time graduation.

While the research literature shows a high degree of consensus on the factors that place students at risk of failure to graduate on time, there is no guarantee that a given indicator will predict failure to graduate on time with the same accuracy if it is applied to students in different school contexts. Most previous studies have focused primarily on large urban centers, and even there, indicators' value as predictors and cutpoints for the indicators

differ across districts. For example, a 2011 Regional Educational Laboratory (REL) Southwest study of five Texas school districts (Hartman, Wilkins, Gregory, Gould, & D'Souza, 2011) and a 2012 REL Midwest study of two urban midwestern districts (Norbury et al., 2012) found that although the on-track indicators were highly accurate predictors of graduation in Chicago Public Schools (Allensworth & Easton, 2005), their accuracy varied considerably when applied to other districts. A review and re-analysis of indicators identified in 36 published articles also reports considerable variation in the accuracy of individual indicators (Bowers, Sprott, & Taff, 2013).

Because the literature suggests that the accuracy of indicators for predicting graduation outcomes may vary by context, researchers advise school districts to independently verify the accuracy of indicators using their own data before applying them in an early warning system (Gleason & Dynarski, 2002). Improving and tailoring a set of indicators may better identify students at risk of failure to graduate on time (Heppen & Therriault, 2008; Jerald, 2006). This study assisted three Ohio school districts with developing locally validated early warning systems. For districts that choose to validate their own indicators, the report lays out a step-by-step process that school district personnel can follow.

The purpose of this study was to develop a set of locally tailored early warning indicators for students at different grade levels in each school district and examine the accuracy of the indicators for predicting failure to graduate on time

What the study examined

The three Ohio districts in this study are implementing early warning systems to identify, provide services to, and track the progress of students in grades 8 and 9 who are at risk of failing to graduate from high school on time. As the literature suggests, indicators that are accurate predictors of graduation outcomes for students in a particular grade within a particular district may not be as accurate for students in other grades or in other districts (Hartman, et al., 2011; Norbury, et al., 2012). Furthermore, the cutpoints that most accurately classify students as on or off track may also differ across districts or for students in different grades within the same district. The purpose of this study was to develop a set of locally tailored early warning indicators for students at different grade levels in each school district and examine the accuracy of the indicators for predicting failure to graduate on time.

Research questions

This study addressed the following research questions:

- For each candidate indicator, what is the optimal cutpoint for accurately classifying students as on track or off track to graduate? How do these cutpoints vary across districts and grades?
- Which indicators consistently predict failure to graduate on time when their optimal cutpoints are used?
- Which of the consistently predictive indicators have the highest correct off-track prediction rates, lowest false alarm rates, and best overall accuracy? How does the accuracy of indicators vary across districts and grades?

Indicators were selected as candidates for validation based on a review of the literature (see appendix A). The indicators are based on three types of information: student attendance, academic achievement, and discipline. For grade 8 students, included data were related to attendance rate, grade point average, number of failing grades (overall and in core courses), reading and math scores on the Ohio Achievement Assessment, and number of suspensions. For grade 9 students, included data were related to attendance

rate, grade point average, number of credits earned, number of failing grades (overall and in core courses), and number of suspensions. The number of indicators included in the analysis differed across the three districts because of differences in the availability and consistency of raw data on the grade 8 or grade 9 cohorts. Definitions of candidate indicators are given in box 2.

Four criteria for identifying valid indicators of failure to graduate on time

The study team reviewed the research literature and generated a list of grade 8 and 9 data elements that were commonly found to be significant predictors of failing to graduate from high school (see literature review in appendix A). The list included student attendance, achievement, coursework, and discipline data elements. Districts were asked to review the list and provide the data elements, if available, for students in the cohorts that were expected to graduate in 2010 and 2011. This resulted in a set of grade 8 and 9 candidate early warning indicators for each of the three districts. A set of four criteria were developed for judging whether particular data elements or sets of elements were valid early warning indicators:

- The indicator must provide early warning of students' risk of failure to graduate on time. Specifically, candidate indicators for this study had to reflect student

A review of the research literature generated a list of grade 8 and 9 data elements that were commonly found to be significant predictors of failing to graduate from high school, including student attendance, achievement, coursework, and discipline data elements

Box 2. Definitions of candidate early warning indicators

End-of-year attendance rate. The proportion of total number of days that a student attended school, had an excused absence, or had in-school suspension to the total number of days that the student was expected to attend school (as specified in Ohio Administrative Code 3301–18–01 of 2008).

Grade point average. Students' average academic achievement in both core and elective courses. Most grade point averages are on a four-point scale. District C's end-of-year grade point average was provided on a five-point scale but was rescaled to a four-point scale for grade 9 students to allow for comparisons with the other districts.

Number of credits earned. The cumulative number of credits a student has earned in both core and elective courses.

Number of failing grades. The cumulative number of failing grades appearing on all quarterly marking periods or semesters over the school year. For Districts A and C the total number of failing grades equals the cumulative number of failing grades appearing on report cards from all four quarterly marking periods. For District B it is the number of failing grades received on two semester report cards (fall and spring).

Number of failing grades in core courses. The cumulative number of failing grades appearing on all quarterly marking periods or semesters over the school year in core courses. For Districts A and C the total number of failing grades equals the cumulative number of failing grades appearing on report cards from all four quarterly marking periods. For District B it is the number of failing grades received on two semester report cards (fall and spring).

Number of suspensions. The number of times a student received a suspension as a disciplinary measure over the school year. This includes both in-school and out-of-school suspensions.

Reading and math scores on the Ohio Achievement Assessment. Student reading and math scores on the Ohio Achievement Assessment in grade 8. Both scores range from 250 to 500, with 400 being the minimum score required for a rating of proficient.

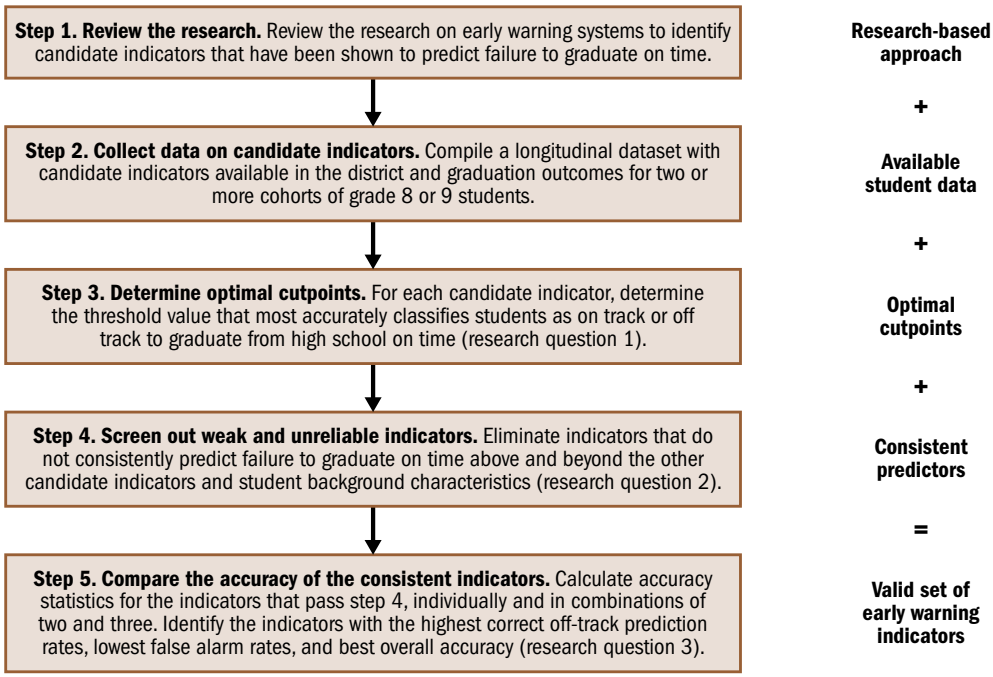
performance or behavior in grades 8 and 9, thus giving educators an opportunity to provide three to four years of support.

- The indicator must be easily communicated to educators within a district. To meet this criterion, the study team identified the value for each candidate indicator that separated students who would most likely graduate on time from those who would not. This value is referred to as the optimal cutpoint.
- The indicator must show a statistically significant relationship with students' four-year graduation outcomes in the two-cohort dataset used in this analysis.
- The indicator must show evidence that it will be a consistent predictor of failure to graduate on time when applied to future cohorts of students from the same district. This requires an indicator to significantly predict failure to graduate in at least 50 of 100 randomly simulated cohorts.

The accuracy of each indicator (or combination of indicators) was assessed on the basis of how well it maximized correct off-track predictions and minimized false alarms

After identifying indicators that meet these criteria, the study team compared their accuracy by examining the percentage of students who failed to graduate on time in 2010 and 2011 who had been correctly identified as off track by the grade 8 or grade 9 indicator (the correct off-track prediction rate) and the percentage of 2010 and 2011 graduates who had been incorrectly flagged as off track (the false alarm rate). The accuracy of each indicator (or combination of indicators) was assessed on the basis of how well it maximized correct off-track predictions and minimized false alarms. The steps in identifying indicators are illustrated in figure 1, and data sources and research methods are given in box 3.

Figure 1. A systematic process for identifying valid early warning indicators of failure to graduate on time



Note: Failure to graduate on time was defined as failure to earn a high school diploma within four years. Steps 3–5 correspond to the three research questions.

Source: Authors' compilation.

Box 3. Data sources and research methods

The types of student data used to predict failure to graduate in previous studies (see a review of the research literature on early warning systems in appendix A) were requested from the three participating districts. The data elements collected generally included grade 8 and 9 student attendance, achievement, coursework, and discipline records.

Each district provided grade 8 and 9 student data as well as the graduation outcomes for two cohorts of students. The first cohort enrolled in high school in fall 2006 and were expected to graduate in spring 2010. The second cohort enrolled in fall 2007 and were expected to graduate in spring 2011.

To answer the first research question on optimal cutpoints for the indicators, the study team used receiver operating characteristic curve analysis to determine for each candidate indicator the cutpoint that maximizes the correct off-track prediction rate and minimizes the false alarm rate. The optimal cutpoint is determined by calculating the area-under-the-curve statistic, which is referred to as overall accuracy. An area-under-the-curve statistic of less than .50 indicates that the variable is no better at predicting which students will not graduate than a random guess; a higher statistic indicates greater accuracy.

To answer the second research question on which data elements were the most consistent indicators of failure to graduate on time, the study team used a statistical technique called stepwise logistic regression to identify the subset of indicators for each district and grade that were statistically significant predictors of student failure to graduate on time above and beyond the other candidate indicators. A validation test was applied to verify that the indicators identified were consistent predictors of failure to graduate on time when applied to variations of the sample of students included in the logistic regression analysis. This technique involved drawing 100 subsamples of students from the datasets for each district and grade and repeating the stepwise logistic regressions on each subsample to see whether the indicators were repeatedly identified as more statistically significant than the other candidate indicators (see Chen & George, 1985; Mick & Ratain, 1994).

For the third research question the study team calculated which consistently predictive early warning indicators had the highest off-track prediction rates, the lowest false alarm rates, and the best overall accuracy (best balance between correct off-track predictions and false alarms).

What the study found

End-of-year attendance rate was the only indicator that consistently predicted graduation outcomes for both grades 8 and 9 in all three districts

Certain indicators were more accurate predictors of failure to graduate on time in some districts than others, and the optimal cutpoints for classifying students as on track or off track for graduation differed across districts. End-of-year attendance rate was the only indicator that consistently predicted graduation outcomes for both grades 8 and 9 in all three districts. Reading scores on the Ohio Achievement Assessment were consistent indicators for grade 8 students in all three districts. The most accurate indicators in both grade 8 and grade 9 were based on coursework (grade point average and number of credits earned). Consistent with prior literature, failing more than one class and being suspended one or more times were also strong predictors of failure to graduate on time. On average, indicators were more accurate in grade 9 than in grade 8.

The findings are presented in detail in the following sections.

Optimal cutpoints for five of the eight indicators varied by district and grade level

The first research question asks about the cutpoints that most accurately separate students who graduate on time from those who do not for each candidate indicator and how these cutpoints vary across districts and grades. The analysis assumed that districts are equally concerned about maximizing the number of students correctly identified as off track and about minimizing the number of students incorrectly identified as off track. The optimal cutpoints for five of the eight candidate indicators varied by district and grade level within districts (table 1). Two grade 9 indicators had the same optimal cutpoint in all three districts: number of suspensions (one or more) and number of failing grades (one or more). The optimal cutpoints for number of credits earned over the school year was also the same (fewer than seven) for the two districts providing data on this indicator.

Two grade 9 indicators had the same optimal cutpoint in all three districts: number of suspensions (one or more) and number of failing grades (one or more)

Cutpoints for attendance were the same for grade 8 and 9 students in Districts B and C but significantly lower for District A. For grade 8 and 9 students in Districts B and C, attendance rates below 95 percent, which equates to missing 9 of the state-required 182 school days, most accurately predicted failure to graduate on time. In District A, grade 8 attendance rates below 93 percent (13 of 182 days missed) and grade 9 attendance rates below 90 percent (18 of 182 days missed) most accurately predicted failure to graduate on time (confidence intervals for the cutpoints are provided in table B2 in appendix B).

Grade point average cutpoints varied across districts, but with one exception fell between 1.9 and 2.3 (roughly C to C+ averages). District C's grade 8 grade point average cutpoint was significantly higher (3.1); however, in step 4 of the analysis (see figure 1) this indicator was found to be unreliable and eliminated as a candidate for District C's grade 8 early warning system.

The optimal cutpoints for grade 8 reading and math scores on the Ohio Achievement Assessment also varied across districts. In District A the cutpoints in both subjects fell just below the state's proficiency cut score of 400 (cutscore for reading was 399; cutscore

Table 1. On five of the eight data elements examined, the optimal cutpoint for designating students as off track to graduate differed for at least one district, 2006/07 and 2007/08 data for 2010 and 2011 graduation outcomes

	District A		District B		District C	
Indicator	Grade 8	Grade 9	Grade 8	Grade 9	Grade 8	Grade 9
End-of-year attendance rate (percent)	< 93	< 90	< 95	< 95	< 95	< 95
Grade point average	—	< 1.9	< 2.3	< 2.1	< 3.1	< 2.2
Number of failing grades overall	≥ 1	≥ 1	≥ 2	≥ 1	≥ 1	≥ 1
Number of failing grades in core courses	≥ 1	≥ 1	≥ 1	≥ 1	—	—
Number of suspensions	≥ 1	≥ 1	≥ 1	≥ 1	—	≥ 1
Math score on the Ohio Achievement Assessment	< 392	na	< 406	na	< 413	na
Reading score on the Ohio Achievement Assessment	< 399	na	< 413	na	< 420	na
Number of credits earned	na	< 7	na	< 7	na	—

na is not applicable for students at the indicated grade level; — is district did not provide student data or student characteristic or data were incomplete or unusable in the analysis.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

for math was 392), while cutpoints for District B were just above the proficiency cut score. District C's cut scores were in the middle of the proficiency range and close to the state average grade 8 scores (421 for reading and 413 for math).

Consistent predictors of failure to graduate from high school on time were identified for each district

Findings for the second research question show that each district has at least two grade 8 and 9 indicators that consistently predict failure to graduate from high school on time.

Tables 2–4 report the consistent indicators for districts A, B, and C for grades 8 and 9 and the three accuracy metrics (correct off-track prediction rate, false alarm rate, overall accuracy) for all indicators identified as consistent. The tables also present accuracy metrics for generic "combination" indicators that classify students as off track if they fall below (or above) the optimal cutpoints on one or more, two or more, or three or more of the consistent indicators identified for a particular grade and district.

Each district has at least two grade 8 and 9 indicators that consistently predict failure to graduate from high school on time

District A had the most indicators qualifying as consistent (see table 2). This result was anticipated because District A's cohorts were larger and had higher rates of failure to graduate on time, which makes it easier to detect significant relationships between the indicators and the failure to graduate on time outcome. All six candidate indicators in District A's grade 8 dataset proved to be consistent predictors of students' graduation outcome four years later. In grade 9, five of the six candidate indicators were consistent predictors for students' graduation outcomes. The only indicator from District A's grade 9 dataset that was eliminated was one or more failing grades in all courses. This indicator was not a significant predictor of failure to graduate on time after the other indicators were controlled for. It was also redundant with the indicator for failing one or more core courses but less reliable. Detailed results of the indicator consistency tests are in appendix B.

Four of the seven candidate indicators from District B's grade 8 dataset were consistent predictors of student failure to graduate on time: grade point average below 2.3, two or more failing grades, reading score on the Ohio Achievement Assessment below 413, and attendance rate below 95 percent (see table 3). Likewise, four of the six candidate indicators for grade 9 students in District B qualified as consistent: fewer than seven credits earned, one or more suspensions, grade point average below 2.1, and attendance rate below 95 percent.

In District C three of the five grade 8 indicators and two of the four grade 9 indicators passed the tests for consistency (see table 4). In grade 8 one or more failing grades across all course subjects, attendance rate below 95 percent, and reading score on the Ohio Achievement Assessment below 420 were consistent predictors of failure to graduate on time. In grade 9 a grade point average below 2.2 and attendance rate below 95 percent were consistent predictors.

The most accurate indicators of failure to graduate on time differed by school district

The third research question asked which consistently predictive early warning indicators had the highest off-track prediction rates, the lowest false alarm rates, and the best overall accuracy (best balance between correct off-track predictions and false alarms). These three metrics were calculated for Districts A, B, and C.

Table 2. District A accuracy of grade 8 and 9 consistent indicators and combination indicators in predicting failure to graduate on time, 2006/07 and 2007/08 data for 2010 and 2011 graduation outcomes

Indicator	Correct off-track prediction rate[a] (percent)	False alarm rate[b] (percent)	Overall accuracy[c] (proportion)
Grade 8			
Consistent indicators			
Attendance rate below 93 percent	63	31	.67
One or more failing grades in all courses[d]	52	21	.65
Reading score on the Ohio Achievement Assessment below 399	63	34	.64
Math score on the Ohio Achievement Assessment below 392	60	35	.62
One or more failing grades in core courses[d]	36	12	.62
One or more suspensions	22	9	.57
Combinations			
Off track on one or more consistent indicators	86	63	.62
Off track on two or more consistent indicators	70	37	.66
Off track on three or more consistent indicators	51	18	.66
Grade 9			
Consistent indicators			
Grade point average below 1.9	67	20	.73
One or more failing grades in core courses[d]	56	16	.70
Attendance rate below 90 percent	65	28	.69
Fewer than seven credits earned	65	30	.68
One or more suspensions	36	15	.60
Combinations			
Off track on one or more consistent indicators	85	58	.64
Off track on two or more consistent indicators	79	46	.66
Off track on three or more consistent indicators	67	26	.71

a. Percentage of nongraduates accurately classified as off track.

b. Percentage of graduates incorrectly classified as off track.

c. The area-under-the-curve statistic, which ranges from .50 to 1.00, with higher values associated with higher accuracy (higher correct off-track prediction rate and lower false alarm rate).

d. Total number of failing grades on four marking-period report cards.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

In District A using a combination indicator that flags all students who meet the off-track criteria on any one of the five consistent grade 8 indicators significantly increases the correct off-track prediction rate to 86 percent

In District A's grade 8 dataset the single indicator that most accurately predicted failure to graduate on time was attendance rate below 93 percent (see table 2). This indicator also had the highest correct off-track prediction rate of all individual indicators, with 63 percent of all nongraduates correctly classified as off track. However, using a combination indicator that flags all students who meet the off-track criteria on any one of the five consistent grade 8 indicators significantly increases the correct off-track prediction rate to 86 percent. The grade 8 indicator with the lowest false alarm rate in District A was one or more suspensions (9 percent).

For grade 9 students, grade point average below 1.9 had the highest overall accuracy for predicting failure to graduate on time. This indicator had a correct off-track prediction rate of 67 percent while keeping false alarms to 20 percent. Consistent with the findings for District A's grade 8 indicators, the correct off-track prediction rates for grade 9 can be

Table 3. District B accuracy of grade 8 and 9 consistent indicators and combination indicators in predicting failure to graduate on time, 2006/07 and 2007/08 data for 2010 and 2011 graduation outcomes

Indicator	Correct off-track prediction rate[a] (percent)	False alarm rate[b] (percent)	Overall accuracy[c] (proportion)
Grade 8			
Consistent indicators			
Grade point average below 2.3	62	23	.70
Two or more failing grades in all courses[d]	59	23	.68
Reading score on the Ohio Achievement Assessment below 413	60	28	.66
Attendance rate below 95 percent	57	29	.64
Combinations			
Off track on one or more consistent indicators	87	62	.63
Off track on two or more consistent indicators	77	43	.67
Off track on three or more consistent indicators	69	29	.70
Grade 9			
Consistent indicators			
Fewer than seven credits earned	74	27	.74
One or more suspensions	45	10	.68
Attendance rate below 95 percent	67	33	.67
Grade point average below 2.1	62	23	.70
Combinations			
Off track on one or more consistent indicators	89	49	.70
Off track on two or more consistent indicators	77	33	.72
Off track on three or more consistent indicators	69	25	.72

a. Percentage of nongraduates accurately classified as off track.

b. Percentage of graduates incorrectly classified as off track.

c. The area-under-the-curve statistic, which ranges from .50 to 1.00, with higher values associated with higher accuracy (higher correct off-track prediction rate and lower false alarm rate).

d. Total number of failing grades on two semester report cards.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

In District B the combination indicator that flags all students who are off track on one or more of the grade 8 indicators increases the correct off-track prediction rate to 87 percent, but it also raises the false alarm rate to 62 percent, giving it a lower overall accuracy than grade point average below 2.3

increased by switching to the combination indicator that captures all students who are off track on one or more consistent predictors. This increases the correct off-track prediction rates to 85 percent but comes at the cost of large increases in the number of students incorrectly flagged as off track; the false alarm rates rise to 58 percent in grade 9. The grade 9 indicator with the lowest false alarm rate is one or more suspensions (15 percent). Despite the low false alarm rate, the suspension indicator has the lowest overall accuracy of all consistent predictors because it flagged only 36 percent of all future nongraduates.

In District B the grade 8 indicator with the best overall accuracy was grade point average below 2.3. This indicator had the highest correct off-track prediction rate (62 percent) and the lowest false alarm rate (23 percent) of the four consistent grade 8 indicators for District B. The combination indicator that flags all students who are off track on one or more of the grade 8 indicators increases the correct off-track prediction rate to 87 percent. However, it also raises the false alarm rate to 62 percent, giving it a lower overall accuracy than grade point average below 2.3.

Table 4. District C accuracy of grade 8 and 9 consistent indicators and combination indicators in predicting failure to graduate on time, 2006/07 and 2007/08 data for 2010 and 2011 graduation outcomes

Candidate indicator	Correct off-track prediction rate[a] (percent)	False alarm rate[b] (percent)	Overall accuracy[c] (proportion)
Grade 8			
Consistent indicators			
Reading score on the Ohio Achievement Assessment below 420	69	26	.72
One or more failing grades in all courses[d]	46	6	.70
Attendance rate below 95 percent	69	30	.69
Combinations			
Off track on one or more consistent indicators	90	57	.66
Off track on two or more consistent indicators	71	33	.69
Off track on all three consistent indicators	54	20	.67
Grade 9			
Consistent indicators			
Grade point average below 2.2	66	15	.75
Attendance rate below 95 percent	70	27	.71
Combinations			
Off track on one or more consistent indicators	91	43	.74
Off track on both consistent indicators	77	21	.78

a. Percentage of nongraduates accurately classified as off track.

b. Percentage of graduates incorrectly classified as off track.

c. The area-under-the-curve statistic, which ranges from .50 to 1.00, with higher values associated with higher accuracy (higher correct off-track prediction rate and lower false alarm rate).

d. Total number of failing grades on four marking-period report cards.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

In District C the most accurate grade 8 indicator was reading score on the Ohio Achievement Assessment below 420

In District B the grade 9 indicator with the best overall accuracy was earning fewer than seven credits (which was also the second most accurate indicator identified across all districts and grades). This grade 9 indicator had a correct off-track prediction rate of 74 percent (the highest of the four grade 9 indicators) and a false alarm rate of 27 percent. The indicator one or more suspensions had the lowest false alarm rate (10 percent), but as with District A, it greatly underpredicted the number of students truly at risk of failure to graduate on time (the correct off-track prediction rate was 45 percent).

In District C the most accurate grade 8 indicator was reading score on the Ohio Achievement Assessment below 420 (see table 4). At .72, this indicator had the highest overall accuracy of all the grade 8 indicators analyzed across the three districts. It also had the highest correct off-track prediction rate (69 percent). The indicator one or more failing grades in all courses had the lowest false alarm rate in the study; it incorrectly flagged 6 percent of graduates as off track while correctly flagging 46 percent of nongraduates.

Of District C's two consistent predictors for grade 9, grade point average below 2.2 had a higher overall accuracy rate than attendance rate below 95 percent. The attendance indicator had a high correct off-track prediction rate (70 percent), while the grade point average indicator had a low false alarm rate (15 percent). When these two indicators are

used together as a combination indicator (grade point average below 2.2 and attendance rate below 95 percent), the correct off-track prediction rate increases to 77 percent and the false alarm rate remains relatively low at 21 percent, which results in an overall accuracy of .78, higher than any other grade 8 or 9 indicator tested across the three districts.

Implications of the study findings

The findings are intended to help the three districts studied make well-informed choices about which indicators to use in their early warning systems and to provide them with clear expectations about the accuracy of their early warning indicators when applied to current and future cohorts of students. The process of identifying grade 8 and 9 indicators that consistently predict failure to graduate on time can serve as a model for other districts to use when designing their own early warning systems.

The findings underscore why it is valuable for districts to analyze their own data and identify the best indicators for their early warning system or at least to verify that the indicators used are valid for their student population

The indicator cutpoints that most accurately identified students as off track to graduate differed for students in Districts A, B, and C. When the optimal cutpoints were used, some indicators were found to be consistent and accurate predictors of failure to graduate on time in one district but not in the others. These findings underscore why it is valuable for districts to analyze their own data and identify the best indicators for their early warning system or at least to verify that the indicators used are valid for their student population. Applying indicators from a different school district may lead to the wrong students being identified as off track. For example, if District A's 90 percent optimal cutpoint for attendance in grade 9 were applied to grade 9 students in District B instead of District B's optimal cutpoint of 95 percent, it would reduce the number of students who are correctly flagged as off track to graduate from 67 percent to 25 percent.

The number of indicators passing the tests for consistency varied across districts and grades, ranging from two (District C grade 9) to six (District A grade 8). On average, indicators were more accurate in grade 9 than in grade 8. Attendance rate was the only indicator available in all six district- and grade-specific datasets, although optimal cutpoints differed across districts and grades. While grade point average consistently predicts failure to graduate on time in four of five disrict-grade combinations that provided these data, the cutpoint below which students are at a higher risk of not graduating on time varies across districts and grades. One or more failing grades in any course was also a consistent predictor of failure to graduate on time in all but one of the district-grade datasets. This is consistent with findings from prior studies (Allensworth & Easton, 2007; Balfanz et al., 2007).

In the three district- and grade-specific datasets for which one or more suspensions was a consistent predictor, the indicator had the lowest false alarm rate (9–15 percent). This may make the indicator attractive for districts with limited resources that want to target a subgroup of students who are at high risk of failure to graduate on time. However, the drawback of suspension indicators is that they greatly underpredict the number of students who are off track because they miss those who are staying out of trouble but still failing academically. Across all three districts, fewer than 50 percent of nongraduates were correctly designated as off track by these indicators.

Using a combination indicator that flags students who are off track on any one or more indicators captures the most students who are truly off track to graduate, but it comes at the cost of incorrectly flagging a high percentage of students who are not at risk of failure

to graduate on time. This catch-all approach may make sense for districts that desire to intervene with as many students as possible who are truly at risk of failure to graduate on time, even if it means a higher percentage of the students who receive dropout prevention services may not need them. However, districts that want to maximize the accuracy of their early warning systems and target their resources to students who are truly at risk might consider using a combination indicator that requires students to meet the off-track criteria on two or more or three or more indicators, instead of just one. Findings from this study show that across all districts and grades, the two- and three-indicator combinations have higher overall accuracy than one-indicator combinations.

Examining indicators' off-track prediction rates, false alarm rates, and overall accuracy can help districts make informed decisions about the indicators that are the best fit for the goals of their early warning systems and the resources available to support dropout prevention. Because districts operate within budget and human resource constraints, selecting indicators that overpredict the number of students who will not graduate on time will result in an inefficient use of resources. However, if an early warning system's goal is to catch as many at-risk students as possible, districts might want to avoid underprediction and identify as many at-risk students as possible, even if it means including some students who are not actually at risk. Using the indicators with the best overall accuracy will lead to the most efficient use of resources, as those indicators achieve the best balance between correct off-track predictions and false alarms. Choosing a system that balances these values ultimately depends on the local context and the system's intent.

Using a combination indicator that flags students who are off track on any one or more indicators may make sense for districts that desire to intervene with as many students as possible who are truly at risk of failure to graduate on time, even if it means a higher percentage of the students who receive dropout prevention services may not need them

Limitations of the study

The grade 8 and 9 indicators analyzed in this study were limited to those that were consistently identified as predictive of failure to graduate on time in the research literature on early warning systems and were based on data elements that a typical school district routinely collects and stores. Some school districts may collect and store other data elements on grade 8 and 9 students that more accurately predict failure to graduate on time, and these can be identified by following the analytic processes outlined in this report.

This study used only the four-year graduation rate as the outcome because it was of most interest to the members of the Midwest Dropout Prevention Research Alliance and is the accountability measure tracked by the state of Ohio. Some studies examine both four- and five-year graduation rates. This analysis treated students who complete high school in five years as students who did not graduate on time (that is, within four years).

The analysis was restricted to students who were first-time freshmen within the districts in 2006/07 or 2007/08 and excluded students who entered the district after grade 9. Thus, this study does not present a full portrait of the graduating classes of 2010 and 2011. Furthermore, findings based on retrospective cohorts might not apply to current and future cohorts.

Results from the three participating districts, which were not randomly selected, may not generalize to other districts in Ohio or other states. Although the methods presented here for validating local early warning indicators are applicable to other districts, the specific findings on the candidate indicators may not apply to other districts.

The findings apply most directly to the specific cohorts of students within these three districts. Additional rounds of validation would be necessary to provide more confidence in the ability of these indicators to predict failure to graduate on time for other cohorts. Validating the indicators on a recent cohort would confirm that the indicators have out-of-sample validity, meaning that their predictive value holds up when they are applied to data from a cohort that was not included in the original research sample.

Another limitation is that the study team did not have data on student participation in dropout prevention initiatives; therefore, it is possible that students who were effectively served by these interventions were not flagged as off track to graduate on time by the early warning indicators. It is also possible that some students who were flagged as off track to graduate ended up graduating because they participated in a dropout intervention, in which case the false alarm rates reported in this study would be higher than expected if the interventions were not in place. Districts can compare the list of students identified as off track with the list of students participating in dropout interventions to see whether there are discrepancies.

The cutpoint method assumes that districts seek to achieve the best balance between overpredicting and underpredicting failure to graduate on time. In practice, districts may be willing to tolerate a higher false alarm rate if it means they will maximize the correct off-track prediction rate or vice versa, in which case the optimal cutpoints would differ from those reported in this study. These decisions will depend on the goals of their early warning systems.

The cutpoint method assumes that districts seek to achieve the best balance between overpredicting and underpredicting failure to graduate on time. In practice, districts may be willing to tolerate a higher false alarm rate if it means they will maximize the correct off-track prediction rate or vice versa, in which case the optimal cutpoints would differ from those reported in this study

The findings on the accuracy of the candidate indicators are optimistic in that they are based on the same sample used to fit the indicators and determine the cutpoints. The bootstrap resampling methods help identify indicators that are consistently predictive of failure to graduate on time when applied to randomized subsamples of the full two-cohort datasets used in the study. However, the characteristics of the study's two cohorts may differ from those of subsequent cohorts in ways that could affect prediction of failure to graduate on time. The chances that these differences between cohorts produce faulty predictions can be reduced by continuously checking whether the indicators remain consistent for subsequent cohorts of students.

Finally, the relationships between indicators and students' graduation outcomes are correlational. They say nothing about a possible causal connection between early indicators and later outcomes. Early warning indicators only distinguish students who have a low probability of on-time graduation.

Appendix A. Literature review

The intent of an early warning system is to systematically use data to identify students at risk of not graduating from high school so they can then be matched with interventions to help them get on track for graduation (Heppen & Therriault, 2008; Jerald, 2006; Kennelly & Monrad, 2007; Neild et al., 2007; Pinkus, 2008). Recent research on academic and behavioral predictors of failure to graduate on time (Allensworth & Easton, 2005, 2007; Balfanz et al., 2007; Neild & Balfanz, 2006; Silver et al., 2008) indicate that such systems are a promising approach—or a necessary prerequisite—to effectively prevent students from dropping out of school (Dynarski et al., 2008).

A growing number of states, districts, and schools are using early warning systems. According to Data Quality Campaign (2011), 18 states produce early warning indicator reports that tell educators and parents which students are at risk of not graduating from high school, and many more states are poised to create such reports. Although decisionmaking based on early warning system reports occurs primarily in districts and schools, the prevalence of systems for tracking and analyzing early warning data at these levels is less known. Bruce, Bridgeland, Fox, and Balfanz (2011) identified only a handful of districts and schools that are implementing early warning systems (for example, eastern Missouri, Nashville, Chicago, Knoxville, and Philadelphia) and concluded that more research is needed. Bruce et al. (2011) found no other attempts to systematically document the proportion of schools and districts that use early warning system tools or analytic processes.

Some evidence of the prevalence of early warning systems can be gleaned from the National High School Center, which actively supported districts and schools in the implementation of early warning systems and offers free, downloadable early warning system tools and implementation guides designed for use in the middle grades and high school (it recently recorded more than 3,000 downloads of these tools by representatives of states, districts, and schools). The National High School Center's collaborative projects, in partnership with Regional Comprehensive Centers or states, have directly supported early warning system implementation in 67 districts in six states.

Strong foundational research points to the best indicators to identify students who are at risk of not graduating on time. Studies of early warning indicators, conducted primarily in large districts, including Chicago Public Schools, the School District of Philadelphia, the Los Angeles Unified School District, and others, have demonstrated that indicators in middle school and the first year of high school can predict whether students are on a path toward eventual graduation. These indicators are consistently related to attendance; course performance; credit attainment; and, in some cases, performance on state assessments, grade retention, and behavior (Allensworth & Easton, 2005, 2007; Neild & Balfanz, 2006; Roderick, 1993; Silver et al., 2008).

For middle school students, attendance in grades 6 and 8, course grades in reading and math, and behavior such as out-of-school suspensions are predictive of success in high school and eventual graduation (Balfanz, 2009; Balfanz & Herzog, 2005; Neild & Balfanz, 2006). For example, Balfanz (2009) found that Philadelphia students in grade 6 who failed English language arts or math, attended school less than 80 percent of the time, or received an unsatisfactory behavior grade in a core course had only a 10–20 percent chance of graduating on time. Furthermore, fewer than one out of four students demonstrating at

least one of these indicators graduated from high school within five years. Attendance and academic indicators and thresholds based on this research are commonly used as elements in early warning systems. (See also Heppen & Therriault, 2008, and Jerald, 2006, for summaries of middle school indicators.)

Although early-as-possible intervention is desirable, some of the clearest predictors of the likelihood of graduating are not evident until the first year of high school. In research on rates of students who are on track for graduation in the Chicago Public Schools, Allensworth and Easton (2005, 2007) showed that the most powerful indicators are those related to student engagement (attendance) and course performance (grades and credit accumulation) during the freshman year.

This body of research provides a strong basis for guiding states, districts, and schools to use readily available data about attendance, course performance, and (where available) behavior to create early warning systems. These indicators allow educators to systematically flag students who are missing significant amounts of instructional time, are failing courses, and have behavior issues such as suspensions.

Although the basis is assuredly strong, the cross-context generalizability of this research is not yet known, especially given that most studies have focused on large urban centers, and even those studies found variations. For instance, a 2011 Regional Educational Laboratory (REL) Southwest study of five Texas school districts (Hartman et al., 2011) and a 2012 REL Midwest study of two midwestern districts (Norbury et al., 2012) found that although an on-track indicator (Allensworth & Easton, 2005) predicted on-time graduation in all districts, it varied considerably in predictive strength across districts and was not as strongly predictive as it was in the Chicago Public Schools. These findings suggest that the applicability of the same indicator varies in different contexts. In addition, a REL Mid-Atlantic study on Delaware (Uekawa, Merola, Fernandez, & Porowski, 2010) found that the strongest grade 9 predictors of failure to graduate on time were attendance rate, repeating a grade, and English language arts and math course grades. All three REL reports, as well as the validation investigations by members of the current study team using data from the San Diego Unified School District, the U.S. Virgin Islands, and the state of Massachusetts, underscore the importance of local validation of early warning indicators.

Local validation of indicators, where desired and feasible, focuses on the raw-data predictors that are most strongly related to graduation outcomes, as well as the thresholds applied to the raw-data predictors to turn them into actionable indicators. In some districts, thresholds that work well elsewhere (such as those in the default settings of the National High School Center's early warning tools and similar tools) may overidentify students, leading to long lists of at-risk students that may prove too overwhelming to be useful or underidentify students, leaving some without assistance.

Appendix B. Student samples, data elements, and methodology

This appendix describes the study methodology in detail. A conceptual description of the methodology used for this project is shown in figure 1 in the main text. The study used data on student attendance, achievement, coursework, and discipline for two cohorts of grade 8 and 9 students in three Ohio school districts.

Sample restrictions

Students were excluded from the analysis for two reasons. Because indicators were calculated for students at the end of either grade 8 or 9, students who were not enrolled in the district at the end of grade 8 were excluded from the grade 8 analysis, and students who were not enrolled in the district at the end of grade 9 were excluded from the grade 9 analysis. Students in the grade 8 cohort who were held back (repeated grade 8) were included in the grade 8 sample but excluded from the grade 9 analysis. Second, students were excluded from the analysis if they transferred, left the country, or died in grades 9–12 (for the grade 8 analysis) or grades 10–12 (for the grade 9 analysis). (See the definition of *failure to graduate on time* in box 1 of the main text.) The analysis was restricted to students who were first-time freshmen in the districts in 2006/07 or 2007/08 and excluded students who entered the district after grade 9. Students in the 2006/07 cohort graduated in 2010, and students in the 2007/08 cohort graduated in 2011.

Furthermore, analysis of indicators for district- and grade-specific samples were hampered due to the variation in districts' capacity to collect, maintain, and share data for two cohorts of grade 8 and 9 students (table B1).

Table B1. Districts' capacity to collect, maintain, and share student data on candidate indicators, 2006/07 and 2007/08 data for 2010 and 2011 graduation outcomes

Indicator	District A		District B		District C	
	Grade 8	Grade 9	Grade 8	Grade 9	Grade 8	Grade 9
End-of-year attendance rate	●	●	●	●	●	●
Grade point average[a]	—	●	●	●	●	●
Number of failing grades overall[b]	●	●	●	●	●	●
Number of failing grades in core subjects	●	●	●	●	—	—
Number of suspensions	●	●	●	●	—	●
Reading score on the Ohio Achievement Assessment	●	na	●	na	●	na
Math score on the Ohio Achievement Assessment	●	na	●	na	●	na
Number of credits earned	na	●	na	●	na	—

● indicates that a district provided the student data on that indicator; na is not applicable for students at the indicated grade level; — is district did not provide student data or student characteristic or data were incomplete or unusable in the analysis.

a. District C's end-of-year grade point average is on a five-point scale for grade 9 students, but it was rescaled to a four-point scale to match the other districts.

b. For Districts A and B the number of failing grades reflects the cumulative number of failing grades appearing on report cards from all marking periods during the grade 8 or grade 9 year; for District C it reflects the number of failing grades received on fall and spring semester report cards.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

Description of candidate indicators

Indicators were selected as candidates for validation as early warning indicators based on a thorough review of the literature (see appendix A). As the independent variables, these indicators are based on three types of information: student attendance, academic achievement, and discipline. For grade 9 students, included data were related to attendance rate, grade point average, number of credits earned, number of failing grades (overall and in core courses), number of suspensions, and age. For grade 8 students, included data were related to attendance rate, grade point average, number of failing grades (overall and in core courses), reading and math scores on the Ohio Achievement Assessment, number of suspensions, and age. Definitions of candidate indicators are given in box 2 in the main text.

Validation of the indicators was done separately by grade for each district using unique datasets (one grade 8 and one grade 9 dataset per district), and therefore the study team used all indicators that were available from the districts, regardless of whether the same indicators were available from the other districts or for both grades in the same district. For an indicator to be included in the candidate set of a particular grade and district it could not have missing records for more than 5 percent of enrolled students in either cohort (see table B1).

The number of indicators included in the analysis differed across the three districts because of differences in the availability and consistency of raw data on the grade 8 or grade 9 grade cohorts. All three districts provided the necessary information to calculate the total number of failing grades received by students over the school year for both cohorts and both grades, but for Districts A and C this indicator was based on the cumulative number of failing grades students reported on their four marking-period report cards, whereas in District B it was based on the number of failing grades on two semester report cards.

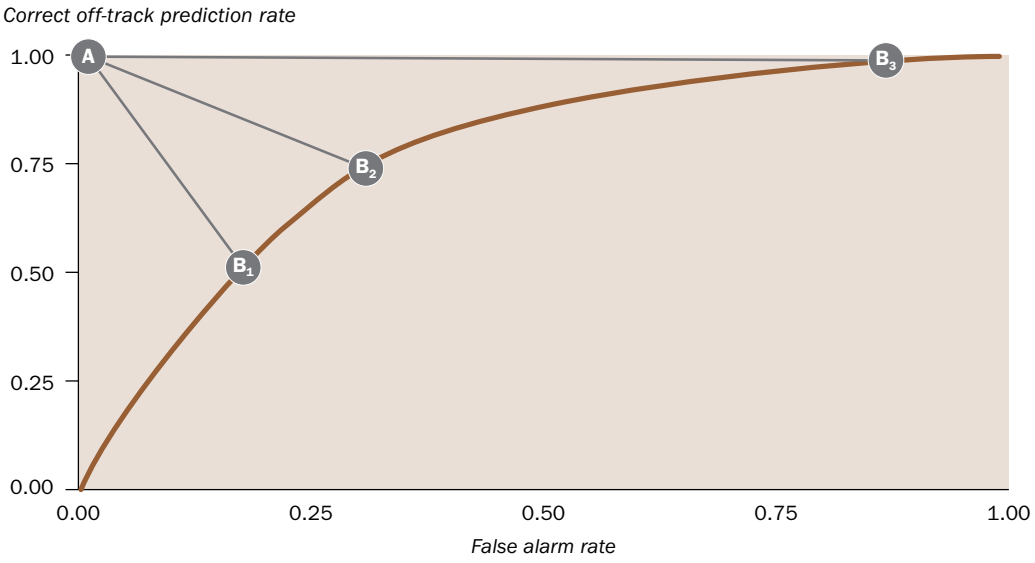
In addition to examining the individual indicators, the study team also examined the performance of combination indicators. A combination indicator is one that classifies students as off track if their performance falls below the optimal cutpoints of any one or more, two or more, or three or more of the consistent indicators.

Analytic strategy for identifying optimal cutpoints (research question 1)

Receiver operating characteristic (ROC) curve analysis was used to determine for each candidate indicator the cutpoint that maximizes the correct off-track prediction rate and minimizes the false alarm rate (Gönen, 2007).

An ROC curve is plotted in an *x*-*y* coordinate system in which the false alarm rate is on the *x*-axis, and the correct off-track prediction rate is on the *y*-axis (figure B1). An ROC curve for a single continuous predictor (independent of any other predictors) plots all possible cutpoints for that predictor. The coordinates for each point in the curve are the false alarm rate and the correct off-track prediction rate associated with that cutpoint. For example, in figure B1, point B1 represents a cutpoint at which the correct prediction rate is .50 and the false alarm rate is .20. The cutpoint for a perfect indicator would have a 100 percent off-track prediction rate and a 0 percent false alarm rate and would be located at point A (0,1). The optimal cutpoint for an indicator is the point in the curve with the shortest distance from point A. In figure B1, point B2 is closer to point A than point B1 or point B3 and would therefore be considered the optimal cutpoint.

Figure B1. Example of a receiver operating characteristic curve



Source: Based on a receiver operating characteristic curve analysis from a report on Delaware indicators and indicator thresholds (Uekawa, Merola, Fernandez, & Porowski, 2010).

The optimal cutpoint is determined by calculating the area-under-the-curve (AUC) statistic, which is referred to as "overall accuracy" in tables 2–4 in the main text. The AUC statistic is a measure of the balance of the correct off-track prediction rate and the false alarm rate. An AUC statistic of less than .5 indicates the variable is not effective at separating graduates from nongraduates; that is, it is no better at predicting a graduation outcome than a random guess (Fawcett, 2006). As an AUC statistic increases, the correct off-track prediction rate improves and the false alarm rate declines. Therefore, the larger the AUC statistic, the better the indicator classifies nongraduates as nongraduates and does not classify graduates as nongraduates. Cutpoints and their confidence intervals for candidate indicators are shown in table B2. Guidelines for rounding cutpoints to meaningful values are shown in table B3.

Analytic strategy for identifying consistent indicators (research question 2)

To identify consistent indicators the study team used logistic regression analysis. The first step in this procedure was to run a logistic regression model that regresses a binary indicator of failure to graduate on time on the full set of candidate indicators (in binary form using their optimal cutpoints) for a given district and grade. After running the first model the least statistically significant indicator is dropped and the model is rerun. The fit of the second model then is compared with the first model to determine whether the exclusion of that indicator significantly decreases the model's ability to predict four-year graduation above and beyond the set of covariates. The likelihood ratio chi-square test is used to estimate the statistical significance of the difference between the first and second models. If this test statistic is significantly different from zero ($p < .05$), it signals that the indicator has incremental predictive value, meaning it provides new information on graduation above and beyond the set of student background characteristics. This procedure is repeated until all remaining indicators in the model are found statistically significant, meaning they each provide new information that improves the model's ability to predict failure to graduate on time.

Table B2. Optimal cutpoints with confidence intervals for candidate early warning indicators, 2006/07 and 2007/08 data for 2010 and 2011 graduation outcomes

Indicator	District A		District B		District C	
	Grade 8	Grade 9	Grade 8	Grade 9	Grade 8	Grade 9
End-of-year attendance rate (percent)						
Cutpoint	< 93	< 90	< 95	< 95	< 95	< 95
95 percent confidence interval	± 0.01	± 0.01	± 0.05	± 0.05	± 0.07	± 0.06
Grade point average						
Cutpoint	—	< 2.0	< 2.3	< 2.1	< 3.1	< 2.2
95 percent confidence interval	—	± 0.06	± 0.15	± 0.12	± 0.65	± 0.21
Number of failing grades overall						
Cutpoint	≥1	≥1	≥2	≥1	≥1	≥1
95 percent confidence interval	± 0.0	± 0.0	± 0.0	± 0.0	± 0.0	± 0.0
Number of failing grades in core courses						
Cutpoint	≥1	≥1	≥1	≥1	—	—
95 percent confidence interval	± 0.0	± 0.0	± 0.0	± 0.0	—	—
Number of suspensions						
Cutpoint	≥1	≥1	≥1	≥1	—	≥1
95 percent confidence interval	± 0.0	± 0.0	± 0.0	± 0.0	—	± 0.0
Reading score on the Ohio Achievement Assessment						
Cutpoint	399	na	413	na	420	na
95 percent confidence interval	± 1.2	na	± 1.8	na	± 2.1	na
Math score on the Ohio Achievement Assessment						
Cutpoint	392	na	406	na	413	na
95 percent confidence interval	± 1.3	na	± 1.7	na	± 2.3	na
Number of credits earned						
Cutpoint	na	< 7	na	< 7	na	—
95 percent confidence interval	na	± 0.0	na	± 0.0	na	—

na is not applicable to students at the indicated grade level; — is district did not provide student data or student characteristic or data were incomplete or unusable in the analysis.

Note: The 95 percent confidence interval is around the cutpoint value. Optimal cutpoints are rounded to meaningful values based on the guidelines in table B3.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

Model specification. The outcome variable Y will take the value 1 if a student does not graduate from high school within four years, and 0 otherwise. To model this binary outcome, the probability of graduating from high school on time, $\phi_{ij} = P(Y_{ij} = 1)$ is transformed using the logit link function, $\eta = \text{Log}[\phi/(1 - \phi)]$. The transformed variable will then be modeled as the outcome in the regression models taking the form:

$$\eta = \beta_0 + \beta_{1i} p_i + \beta_3 c \qquad \text{(B1)}$$

where β_0 is the estimated log-odds of failure to graduate on time for a student who is not flagged by any indicators, p_i is the ith indicator from the set of candidate indicators in table 1 in the main text, which is equal to 1 if the student is flagged by the indicator, and c is a cohort indicator.

Table B3. Rounding guidelines for creating locally specific indicators based on optimal cutpoints derived from receiver operating characteristic curve analysis

Indicator	Rounded cutpoint value
End-of-year attendance rate (percent)	Whole number
Grade point average	One decimal place
Total number of failing grades overall	Whole number
Total number of failing grades in core subjects	Whole number
Reading score on the Ohio Achievement Assessment	Whole number
Math score on the Ohio Achievement Assessment	Whole number
Number of credits earned	Whole number

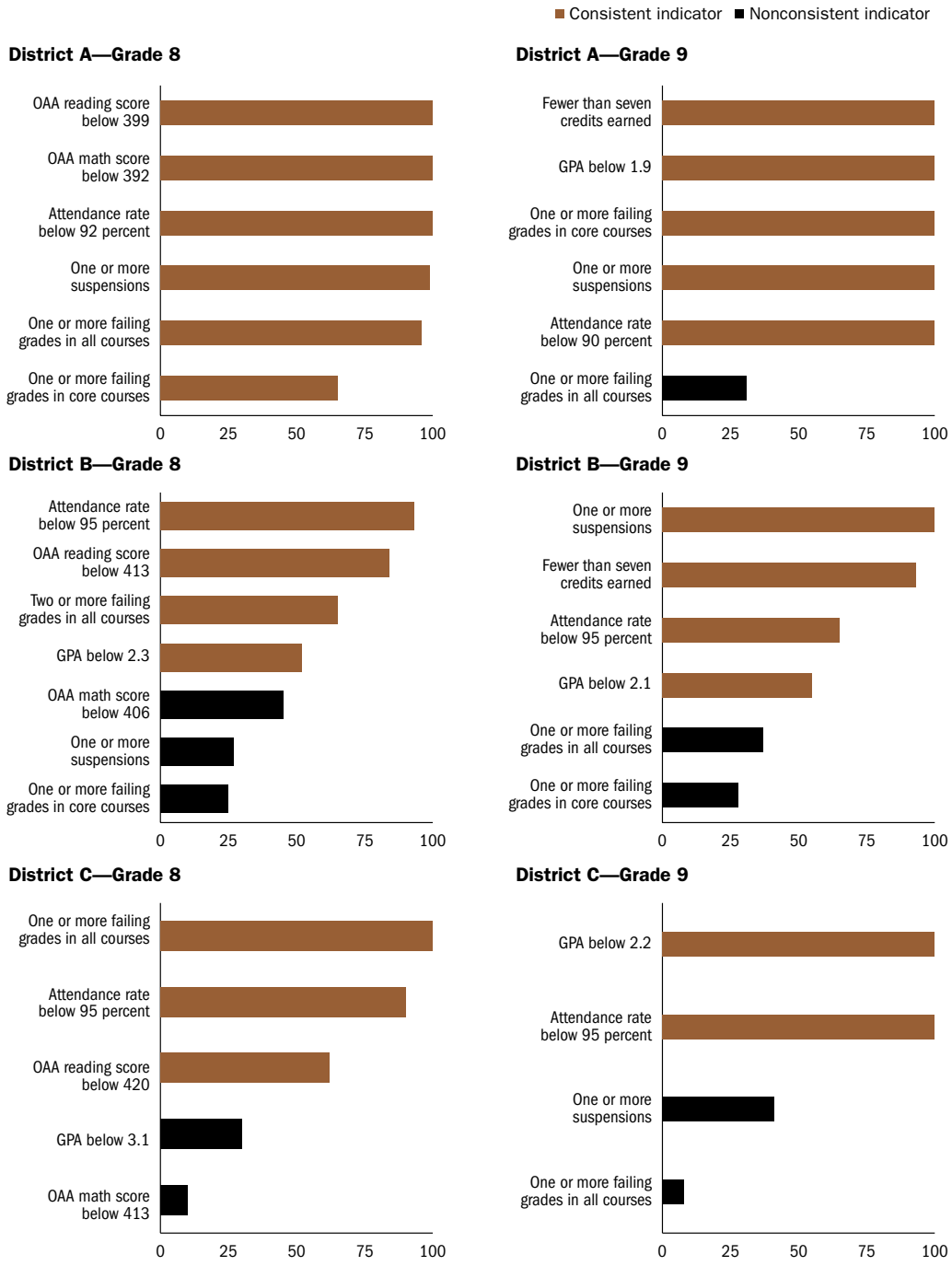
Note. Optimal cutpoints are rounded in the direction nearest to the (1,0) coordinates of the receiver operating characteristic graph.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

The logistic regression models leveraged a statistical technique referred to as bootstrapped resampling. This technique is designed to test whether the predictive value of each indicator, as judged by the statistical test described above, is consistent and not just a function of the particular relationships observed for the particular students in the dataset. Consistent indicators are expected to perform better when applied to new data from their target underlying populations (for example, current and future grade 8 or 9 students). To simulate how an indicator will perform on data from other cohorts, the study team drew 100 random samples from the full two-cohort dataset for a given district and grade; each sample was equal in size to the average of the two cohorts. Each random sample was drawn independently from the full two-cohort dataset, so students could be represented in more than one sample.

The logistic regression described above is conducted for each of the 100 simulated samples. To be identified as a consistent indicator, an indicator must improve the predictive power of the logistic regression model in more than 50 of the randomly simulated cohorts (Chen & George, 1985; Mick & Ratain, 1994; Scalon, Freire, & Cunha, 1998). The test also eliminates indicators that may be strong predictors for the particular cohorts included in the district's dataset but are not expected to consistently predict failure to graduate on time when applied to future cohorts with different characteristics. Results for the tests of consistency are shown in figure B2.

Figure B2. Consistent indicators are statistically significant predictors of failure to graduate on time above and beyond other candidate indicators in at least 50 of 100 randomly simulated cohorts, 2006/07 and 2007/08 data for 2010 and 2011 graduation outcomes



GPA is grade point average; OAA is Ohio Achievement Assessment.

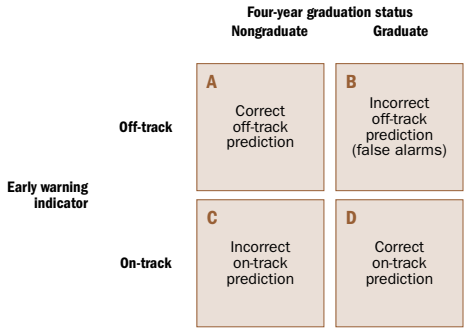
Note: These graphs show the results of the consistency tests for the candidate indicators in each district and grade. The horizontal axis on each chart indicates the number of randomly simulated cohorts for which each candidate indicator was identified as a statistically significant predictor in the stepwise logistic regression models. Indicators that were identified as statistically significant in at least 50 of 100 simulated cohorts were classified as consistent.

Source: Authors' analysis of data on student attendance, achievement, coursework, and discipline provided by the three school districts in the study.

Calculating the correct off-track prediction rate and false alarm rate (research question 3)

The accuracy of each consistent indicator was assessed using the correct off-track prediction rate, the false alarm rate, and the AUC statistic previously described. The correct off-track prediction rate and false alarm rate are derived from combinations of the four possible outcomes shown in the confusion matrix in figure B3. Cells A and D of the matrix indicate correct predictions, and cells B and C indicate the incorrect, or "confusing," predictions.

Figure B3. Confusion matrix and the two metrics used to evaluate the performance of the early warning indicators



Metric	Formula	Interpretation
Correct off-track prediction rate =	$\frac{A}{(A + C)}$	Proportion of nongraduates classified as off track
False alarm rate =	$\frac{B}{(B + D)}$	Proportion of graduates incorrectly classified as off track

Source: Authors' creation.

References

Allensworth, E., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. Chicago, IL: Consortium on Chicago School Research. Retrieved February 20, 2015, from http://ccsr.uchicago.edu/sites/default/files/publications/p78.pdf.

Allensworth, E., & Easton, J. Q. (2007). *What matters for staying on-track and graduating in Chicago public high schools: A close look at course grades, failures, and attendance in the freshman year.* Chicago, IL: Consortium on Chicago School Research. Retrieved February 20, 2015, from http://ccsr.uchicago.edu/sites/default/files/publications/07%20What%20Matters%20Final.pdf.

Balfanz, R. (2009). *Putting middle grades students on the graduation path: A policy and practice brief.* Baltimore, MD: Johns Hopkins University, Everyone Graduates Center & Talent Development Middle Grades Program.

Balfanz, R., & Herzog, L. (2005, March). *Keeping middle grade students on-track to graduation: Initial analysis and implications*. Presentation at the second Regional Middle Grades Symposium, Philadelphia, PA.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grade schools: Early identification and effective interventions. *Educational Psychologist, 42*(4), 223–235. http://eric.ed.gov/?id=EJ780922

Bowers, A. J., Sprott, R., Taff, S. A. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal, 96*(2), 77–100. http://eric.ed.gov/?id=EJ995291

Bruce, M., Bridgeland, J. M., Fox, J. H., & Balfanz, R. (2011). *The use of early warning indicator and intervention systems to build a grad nation*. Washington, DC: Civic Enterprises. Retrieved February 20, 2015, from http://www.civicenterprises.net/MediaLibrary/Docs/on_track_for_success.pdf.

Chen, C., & George, S. (1985). The bootstrap and identification of prognostic factors via Cox's proporptional hazards regression model. *Statistics in Medicine*, 4(1), 39–46.

Data Quality Campaign. (2011). *Hot topic: Supporting early warning systems*. Washington, DC: Author. Retrieved February 20, 2015, from http://www.dataqualitycampaign.org/files/Hot%20Topic_Early%20Warning.pdf.

Dynarski, M., Clarke, L., Cobb, B., Finn, J., Rumberger, R., & Smink, J. (2008). *Dropout prevention: A practice guide* (NCEE 2008–4025). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED502502

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874.

Gleason, P., & Dynarski, M. (2002). Do we know whom to serve? Issues in using risk factors to identify dropouts. *Journal of Education for Students Placed at Risk, 7*(1), 25–41. http://eric.ed.gov/?id=EJ640277

Gönen, M. (2007). *Analyzing receiver operating characteristic curves with SAS*. Cary, NC: SAS Institute.

Hartman, J., Wilkins, C., Gregory, L., Gould, L. F., & D'Souza, S. (2011). *Applying an on-track indicator for high school graduation: Adapting the Consortium on Chicago School Research indicator for five Texas districts* (REL 2011–100). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. http://eric.ed.gov/?id=ED514377

Heppen, J. B., & Therriault, S. B. (2008). *Developing early warning systems to identify potential high school dropouts* (Issue Brief). Washington, DC: American Institutes for Research, National High School Center. Retrieved February 20, 2015, from http://www.betterhighschools.org/pubs/documents/IssueBrief_EarlyWarningSystemsGuide.pdf.

Jerald, C. D. (2006). *Identifying potential dropouts: Key lessons for building an early warning data system*. Washington, DC: Achieve. http://eric.ed.gov/?id=ED499838

Kennelly, L., & Monrad, M. (2007). *Approaches to dropout prevention: Heeding early warning signs with appropriate interventions*. Washington, DC: American Institutes for Research, National High School Center. http://eric.ed.gov/?id=ED499009

Mick, R., & Ratain, M. J. (1994). Bootstrap validation pharmacodynamic models defined via stepwise linear regression. *Clinical Pharmacology and Therapeutics, 56*(2), 217–222.

Neild, R. C., & Balfanz, R. (2006). An extreme degree of difficulty: The educational demographics of the urban neighborhood high school. *Journal of Education for Students Placed at Risk, 11*(2), 123–141. http://eric.ed.gov/?id=EJ736311

Neild, R. C., Balfanz, R., & Herzog, L. (2007). An early warning system. *Educational Leadership 65*(2), 28–33.

Norbury, H., Wong, M., Wan, M., Reese, K., Dhillon, S., & Gerdeman, R. D. (2012). *Using the freshman on-track indicator to predict graduation in two urban districts in the Midwest Region* (REL 2012–134). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. http://eric.ed.gov/?id=ED531421

Pinkus, L. (2008). *Using early-warning data to improve graduation rates: Closing cracks in the education system* (Policy Brief). Washington, DC: Alliance for Excellent Education. http://eric.ed.gov/?id=ED510882

Roderick, M. (1993). *The path to dropping out: Evidence for intervention*. Westport, CT: Auburn House.

Scalon, J., Freire, D., & Cunha, T. (1998). Validation of models for predicting the use of health technologies. *Medial Decision Making,* 8(3), 311–319.

Silver, D., Saunders, M., & Zarate, E. (2008). *What factors predict high school graduation in the Los Angeles Unified School District (*California Dropout Research Project Report 14). Retrieved February 20, 2015, from http://cdrp.ucsb.edu/dropouts/researchreport14.pdf.

Uekawa, K., Merola, S., Fernandez, F., & Porowski, A. (2010). *Creating an early warning system: Predictors of dropout in Delaware* (Technical Assistance Brief). Calverton, MD: Regional Educational Laboratory Mid-Atlantic. Retrieved February 20, 2015, from http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/91/MA1275TAFINAL508.pdf.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research