Applied Research Methods

February 2015

How methodology decisions affect the variability of schools identified as beating the odds

Yasuyo Abe
Phyllis Weinstock
Vincent Chan
IMPAQ International

Coby Meyers
R. Dean Gerdeman
W. Christopher Brandt
American Institutes for Research

Overview

Methodology decisions can affect which schools are identified as "beating the odds"—that is, performing better than expected given the populations they serve. Using data from Michigan, this study demonstrates how the identification of schools changes when statistical methods and technical specifications change. The methodology choices made in identifying beating-the-odds schools are policy decisions that require careful consideration.

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

REL 2015–071

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.

February 2015

This report was prepared for the Institute of Education Sciences (IES) under Contract 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.

Abe, Y., Weinstock, P., Chan, V., Meyers, C., Gerdeman, R. D., & Brandt, W. C. (2015). *How methodology decisions affect the variability of schools identified as beating the odds* (REL 2015–071). 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 number of states and school districts have identified schools that perform better than expected, given the populations they serve, in order to recognize school performance or to learn from local school practices and policies. These schools have been labeled "beating the odds," "high-performing/high-poverty," "high-flying," and other terms that reflect their demonstration of higher academic achievement than schools with similar student demographic characteristics.

If administrators are to learn from these schools, it is important to correctly identify the schools that perform above expectations. However, there is no one right approach to identifying these schools. Typical identification approaches often consider many factors, including policy priorities, available data, resources and capacity (including technical analysis), and stakeholders' preferences. These choices can affect which schools are identified and labeled as exceeding performance expectations.

This report considers the Michigan Department of Education's approach to identifying beating-the-odds schools by using two statistical methods. The first method, the prediction method, identifies a school as beating the odds if it outperforms its predicted level of performance given school demographics by comparing the predicted performance of each school to its actual performance. The second method, the comparison method, identifies a school as beating the odds if it outperforms other demographically similar schools by comparing the performance of each school to the performance levels of the 29 demographically most similar schools in the state.

This report uses Michigan's approach as an example to demonstrate how the choice of statistical methods and technical specifications can change which schools are identified as beating the odds. Michigan's two statistical models produced different results: the comparison method identified fewer than half as many as the prediction method (28 versus 75), with a 39 percent agreement rate. When a change was made to the school performance measure, school characteristic indicator, or school sample configuration, the schools identified as beating the odds changed by varying degrees, with changes in school performance measures causing the biggest difference. Identification results also varied across time. For year-to-year variation from school year 2007/08 through 2010/11, the agreement rate between one year and the next was, on average, less than 50 percent.

The findings confirm the importance of carefully considering the conceptual criteria and technical specifications and measures to be used in identifying schools exceeding performance expectations. Different policy and technical choices may lead to wide variations in resulting lists of schools labeled as beating the odds.

Contents

Why this study?

Some schools, including some high-poverty schools, outperform others with similar student demographic and socioeconomic characteristics. Such schools hold promise because they suggest that academic success can be achieved in challenging school environments. Because policymakers, researchers, and practitioners want to learn from these schools about what works, states and districts generate lists of these high-performing schools to study. Yet the schools on the lists may reflect choices of statistical methods, suggesting that any lessons must be interpreted with caution.

This study examines the technical approaches to identifying these schools and explores and compares the implications of using different student performance measures, demographic characteristics, and school sample configurations, as well as different statistical methods and time periods. It offers education policymakers, state and local education agencies, and researchers issues to consider when developing or reviewing an approach to identifying schools that beat the odds.

This study offers education policymakers, state and local education agencies, and researchers issues to consider when developing or reviewing an approach to identifying schools that beat the odds

Nationwide interest in identifying schools that beat the odds is spurred by efforts to recognize and improve school performance

Educators, administrators, and researchers continue to learn how to better identify performance problems and to identify and implement strategies to support continuous improvement and school turnaround (for example, see Herman et al., 2008; Sebring, Allensworth, Bryk, Easton, & Luppescu, 2006). Identifying and examining schools that exceed achievement expectations given their student demographics—sometimes referred to as "high-flying schools" or schools that "beat the odds"[1]—are part of such efforts. Some state and local education agencies identify beating-the-odds schools to recognize them with awards and to motivate similar schools, especially schools disproportionately serving high-needs students that exhibit lower performance. These higher-performing schools may be studied to identify practices associated with success and to develop or identify effective strategies and interventions for supporting and transforming low-performing schools. For example, Arizona (Waits et al., 2006), Delaware (Grusenmeyer, Fifield, Murphy, Nian, & Qian, 2010), and New York City (Connell, 1999) have processes to identify and learn from beating-the-odds schools.

A few regional and national research studies have explored the processes used to identify beating-the-odds schools. These studies enumerate the factors and approaches used to identify schools, including school population, the time period to be analyzed, the stringency of the performance criteria, and the stability of the performance measures used as indicators of success (see appendix A). This study offers practical considerations for educators, administrators, and researchers in establishing criteria and technical approaches to identify beating-the-odds schools.

Methodology decisions can lead to different results when identifying beating-the-odds schools

Like other state education agencies, the Michigan Department of Education wanted to identify unique policies and practices that distinguish beating-the-odds schools from their counterparts and facilitate the transfer of some of these policies and practices to struggling schools. The Regional Educational Laboratory (REL) Midwest worked with the

department, research staff from two intermediate school districts, and one nonprofit organization in a research alliance focused on improving the state's approach to identifying and learning about beating-the-odds schools.

Michigan's identification approach used two statistical methods. The department noticed that in school years 2009/10 and 2010/11, fewer than a third of schools identified by either method were identified by both methods. And within each method, the list of schools changed from one year to the next, with less than half the schools consistently identified as beating the odds for two consecutive years.[2] In fact, of 184 schools identified by one of the two methods, only 4 were identified by both methods in both 2009/10 and 2010/11.

Although some variation across the two methods was expected in the schools identified as beating the odds, the variation was larger than expected. This raised concerns among research alliance members about the consistency of results and whether these methods can adequately identify schools that exceed performance expectations over time. Schools identified based on a single-year spike in achievement or common year-to-year fluctuations in achievement might be less likely to yield useful lessons on practices associated with school improvement.

The two statistical methods used in Michigan's approach to identifying schools that are beating the odds resulted in more variation than expected, raising concerns about whether these methods can adequately identify schools that exceed performance expectations over time

Research alliance members expressed interest in reviewing the state's technical approach to identifying beating-the-odds schools. In response, REL Midwest examined the state's approach, documenting how the identification results can vary because of decisions about statistical methods, technical specifications (for example, performance measures, school sample configuration, and school characteristics), and time periods examined. Although this study focused on the needs identified by a specific research alliance, the challenges of identifying beating-the-odds schools are not unique to one state. Thus, the study may provide information that can assist other states and districts in developing or revising their technical approaches to identifying schools that exceed performance expectations and in understanding the potential limitations as well as the policy or practical implications of the choices they make.

What the study examined

The Michigan Department of Education's approach to identifying schools performing better than expected was examined to see how the choice of statistical methods and technical specifications can change which schools are labeled as beating the odds.

Using Michigan school and student data, this study first investigated the two statistical methods. One method identified a school as beating the odds if it outperformed its predicted level of performance given school demographics. The other method identified a school as beating the odds if it outperformed other demographically similar schools (box 1). The study then looked at differing technical specifications and changes over time.

The investigations were guided by the following questions:

1. How do the schools identified as beating the odds using the prediction method (schools outperforming their predicted performance) vary from those identified using the comparison method (schools outperforming other demographically similar schools) for a

Box 1. Key features of two methods used to identify beating-the-odds schools in Michigan

The two statistical methods used by Michigan to identify beating-the-odds schools aim to select schools that performed better than expected given demographic backgrounds. Both methods have similar components—including a school performance measure, a set of school demographic measures, and specific statistical criteria—to identify schools performing better than similar schools in the sample.

The prediction method uses regression analyses to predict a school's performance based on school demographic characteristics. Each school's predicted performance is then compared with its actual performance. The school is identified as beating the odds if its actual performance is higher than predicted by a statistically significant margin.

The comparison method compares each school's performance to a group of demographically similar schools. A comparison group of the 29 most demographically similar schools in the state is selected for each school in the sample. The school is identified as beating the odds if its performance is higher than that of every comparison school and is statistically significantly higher than the average performance of the comparison group.

The key features of each method are summarized in the table. The prediction and comparison methods differ in many ways, including how school characteristics are taken into account. See appendix B for technical details on these methods.

Key features of the prediction and comparison methods

Component	Prediction method	Comparison method
Brief description of method	Identifies schools that outperform their predicted performance.	Identifies schools that outperform demographically similar schools.
Use of school performance measure	Compares predicted versus actual outcomes on the performance measure.	Compares outcomes on the performance measure among demographically similar schools.
Use of school demographic characteristics	Uses school characteristics, such as percent of English language learner students, percent eligible for free- or reduced-price lunch, and percent racial/ethnic minority, as variables that "predict" the performance measure.	Uses school characteristics, such as percent of English language learner students, percent eligible for free- or reduced-price lunch, and percent racial/ethnic minority, to identify demographically similar schools.
Identification steps	Computes a predicted performance for each school based on the characteristics above and compares it to the actual performance of the school.	Computes the "distance" between a given school and all other schools in the state as a way of measuring how demographically similar they are. Then, for each school, selects the 29 most demographically similar schools in the state, and compares the school's performance to the performance of these demographically similar schools.
Criteria for identifying schools as beating the odds	If the school's actual performance exceeds the predicted performance beyond a level that might otherwise be due to chance, the school is identified as beating the odds.	If the school's actual performance is statistically significantly higher than the average performance of the 29 most similar schools beyond a level that might otherwise be due to chance, the school is identified as beating the odds.

Source: Authors' analysis as described in the report.

given year? (These are between-method, within-year comparisons, or comparisons of identification results between prediction and comparison methods.)

2. How do the schools identified as beating the odds using a given method for a given year vary when alternative performance assessments, school characteristics, and school grade configurations are used? (These are within-method, within-year comparisons, or comparisons of identification results across alternative technical specifications, given a method and a year.)

3. How do the schools identified as beating the odds using a given method vary from year to year? (These are within-method, between-years comparisons, or comparisons of identification results across years, given a method.)

Thus the study first analyzed how the school identification results varied between the two statistical methods for the 2010/11 school year. Second, it examined, for each method, how the identification results varied due to alternative performance measures, sets of school characteristics, and school samples based on school configuration. (Results for other years are available in appendix B.) Last, it investigated how the identification results for each method changed across consecutive school years from 2007/08 through 2010/11. The study team explored the variation in results that follow from specific choices of statistical methods and technical specifications but did not attempt to isolate the causes of the variation.

A change in performance measures could fundamentally alter how success is defined and measured. Similarly, a change in the selection of school characteristics could affect the understanding of results

Two statistical methods represent different ways to identify beating-the-odds schools

Michigan used two statistical methods to identify schools: the prediction method and the comparison method (see box 1). The prediction method defines beating-the-odds schools as schools that exceed the level of academic performance predicted for them based on their demographic characteristics (for example, the percentage of students who are eligible for free or reduced-price lunch). The comparison method defines beating-the-odds schools as schools that outperform a group of demographically similar schools in Michigan on state assessments.

In applying the two statistical methods, three technical specifications were considered

When applying the two methods, Michigan made additional decisions regarding the school performance measures on which schools are compared, school characteristics to be accounted for, and school sample configuration—whether the samples compared all schools or only schools serving similar grade levels. These decisions are referred to as "specifications" in this report.

Each specification choice could affect which schools are identified. For example, a change in performance measures could fundamentally alter how success is defined and measured. Similarly, a change in the selection of school characteristics could affect the understanding of results. For example, if poverty levels are not accounted for when comparing school performance, the results might reflect performance difference caused by socioeconomic factors more than by the quality of education provided. Given the organizational and developmental differences between elementary schools and high schools, identification results could change depending on whether schools are compared across different levels or only within levels.

The performance and school characteristic measures selected were not consistent across methods (for example, English language learner status was used with the comparison method but not with the prediction method) and have been modified over time (for example, the performance measures changed), which makes assessing the source of the variation in the identification results across methods or years challenging. Michigan's choices for the school performance and characteristics measures also changed by year (see appendix B). With respect to the sample configuration, the state compared schools with each other regardless of the grade levels served, rather than restricting comparisons to schools within comparable grade levels. Michigan's specifications for 2010/11 are shown in table 1. The performance data is from the Top-to-Bottom percentile ranking based on a school composite index developed by the state (box 2). Michigan used data from all public schools for which data were available, including magnet and special program schools (box 3).

The performance and school characteristic measures selected were not consistent across methods and have been modified over time, which makes assessing the source of the variation in the identification results across methods or years challenging

Applying baseline and alternative specifications to assess changes in schools identified as beating the odds

The baseline specifications in table 1 are the performance and school characteristic measures and the school sample configuration applied by Michigan in 2010/11 to identify beating-the-odds schools. The school identification results based on baseline specifications are the benchmark with which subsequent analyses based on alternative specifications are compared. Alternative specifications provide examples of other options that could be considered for performance measures, school characteristics, and school sample configuration. They are not necessarily preferred options but address some of the limitations of the baseline specifications. The baseline and alternative specifications are applied to each of the two identification methods in table 1. Additional discussion of the baseline and alternative specifications is in box 2.

Table 1. Baseline and alternative specifications for each statistical method

Specification area	Baseline specifications		Alternative specifications[a]
	Prediction method	Comparison method	
Performance measure	Michigan Top-to-Bottom ranking percentile	Michigan Top-to-Bottom ranking percentile	A composite academic achievement index developed by the study team based on measures created by the Michigan Department of Education.
School characteristics included in analyses	· Percent English language learner students · Percent eligible for free/reduced-price lunch · Percent racial/ethnic minority · Percent with disabilities	· Percent English language learner students · Percent eligible for free/reduced-price lunch · Percent racial/ethnic minority · Percent with disabilities · School configuration · Locale · Total enrollment · Special education center status · State foundation allowance	· Percent English language learner students · Percent eligible for free/reduced-price lunch · Percent racial/ethnic minority · Percent with disabilities · School configuration · Locale · Total enrollment · Magnet school indicator · Percent female
School sample configuration	The sample includes schools serving all grade levels.	The sample includes schools serving all grade levels. Potential comparison schools include schools serving all grade levels.	The sample is separated by grade level (elementary, middle, and high school grades). Beating-the-odds schools are identified separately by grade levels.

Note: These specifications were applied to all years from 2007/08 to 2010/11.

a. Same for the prediction and comparison methods.

Source: Authors' analysis as described in the report.

Box 2. Determining baseline and alternative specifications

The baseline set of specifications mimics Michigan's approach in 2010/11. Alternative specifications that diverge from the baseline specifications are also shown, and then the identification results of the two methods are compared. Specifically, this study explores options with respect to three specification items: the performance measure, the choice of which school characteristics are used, and the school sample configuration used to pool schools across grade levels. The baseline and alternative specifications of these three items are described below (see appendix B for more details).

Performance measures

Baseline (Michigan's current practice). The state currently uses its Top-to-Bottom percentile ranking as a performance measure to identify schools. The ranking is based on a school composite index developed by the state and takes into account the following:

- *Achievement:* grade 3–8 school average achievement, calculated over the most recent two-year period in math and reading, and a two-year average graduation rate, calculated for high schools.
- *Improvement in achievement:* grade 3–8 change in achievement, based on a four-year achievement trend slope in science (tested only in grades 5 and 8), social studies (tested only in grades 6 and 9), and writing (tested only in grades 4 and 7); grade 11 change in achievement, based on a four-year achievement trend slope in math and reading (calculated using a student's grade 8 and 11 scores); and a four-year average annual graduation improvement rate.
- *Achievement gap:* the largest achievement gaps between two subgroups, calculated based on the top 30 percent of students versus the bottom 30 percent of students.
- *Graduation rate:* graduation rate and graduation rate improvement.

This study uses the Top-to-Bottom ranking as the baseline performance measure with 2010/11 data (for more details on the ranking, see https://www.michigan.gov/ttb). The Top-to-Bottom ranking was not available prior to the 2010/11 school year.

Alternative (authors-developed). As an alternative performance measure, the study constructed a composite performance index from student standardized assessment scores based on Michigan state math and reading tests. This composite index provides a common performance measure for all study years. The alternative performance measure is constructed as follows: first, by computing student *z*-scores for each of the core content areas (math and reading) based on the assessment data for a given school year for all students in the state by grade; second, by taking the average of the student *z*-scores in each content area for each school by grade, creating a school content area performance index; and finally, by calculating the overall mean of the average *z*-scores across all content areas and grades for each school for the school year. (See appendix B for more details on the construction of this measure.)

School characteristic measures

Baseline (Michigan's current practice). Michigan includes a different set of school characteristics with each method. For the prediction method, four school demographic indicators are used as predictors. For the comparison method, those indicators are used as well as four additional

(continued)

Box 2. Determining baseline and alternative specifications *(continued)*

indicators to identify groups of demographically similar schools (see table 1 in the main report for a list of school characteristics used).

Alternative (developed by the study team). As an alternative set of school characteristics, the study team identified a set of characteristics to apply to both methods. The study team conducted a series of regressions examining the extent to which any of the school characteristic measures originally applied in either method significantly predicted the alternative performance measures described earlier. The set of school characteristics that were statistically significant for three out of four years was used as the alternative specification. This provided a common set of school characteristics to include with both methods.

School sample configuration

Baseline (Michigan's current practice). The state used a school sample that included all grade levels (elementary, middle, and high school) with both methods; therefore, high schools may be directly compared with elementary schools, middle schools may be directly compared with high schools, and so forth.

Alternative (developed by the study team). As an alternative to the baseline, the study team stratified, or subdivided, the school sample by grade levels (for example, elementary, middle, or high school) and compared schools within the same grade level. If a school served grades across multiple levels (for example, grades 6–12), the school was included in each of the school-level samples representing the grades that it serves. This provided a more direct comparison of organizationally and instructionally similar sets of schools.

Baseline and alternative specifications were used to generate lists of beating-the-odds schools, and these lists were compared to see how they changed with the set of specifications or "model" used. The baseline model (model A) applies the baseline specification to all three specification items—performance measures, school characteristics, and school sample configuration—equivalent to the state's school identification approach in school year 2010/11. A list of beating-the-odds schools based on the baseline model was generated. Then, lists were generated that altered just one of the three specification items at a time (models B, C, and D; table 2). A final list was generated by applying all three alternative specifications at the same time (model E). The checkmarks in table 2 illustrate the performance measure, school characteristics, and school sample configuration choices under each model. Finally, the lists of beating-the-odds schools generated under each alternative model were compared with the baseline model to gauge the extent to which the identification of schools varied by method and year.

Comparing school identification results between prediction and comparison methods

Next the study team generated school identification lists for the prediction and comparison statistical methods using data from model A (baseline) and alternative model E. Using only these two models kept the specifications of the performance measures, school configurations, and school characteristics as comparable and consistent as possible across the two methods for school year 2010/11. The results are presented in the next section. (Results based on the 2007/08 to 2009/10 data are shown in appendix B.)

Box 3. Data sources and sample

The study team used school and student assessment and demographic records from the Michigan Department of Education for K–12 public schools, covering school years 2007/08–2010/11. Student data were aggregated to the school level (for example, individual student English language learner status was aggregated to create a percent English language learner measure at the school level). The study team also used the Common Core of Data from the U.S. Department of Education, National Center for Education Statistics. The complete list and description of school and student academic performance and demographic indicators used in the study are provided in appendix B.

All K–12 public schools in Michigan were intended to be included in the study; that is, the full study sample consists of all K–12 public schools in the state. However, some schools were excluded from the analyses because they were missing student assessment or school demographic data needed for analyses of specific models for identifying beating-the-odds schools. The analysis sample of schools used in each model is thus smaller than the full study sample of schools. Furthermore, because data requirements vary depending on which performance measure and school demographic indicators were used, the analysis samples vary by method and by model. The total number of public schools in Michigan and the total number of schools included in at least one of the identification models analyzed in the study are shown in table 3. (See appendix B for the number of schools by year and additional discussion on the model-specific analysis sample size).

Although the Michigan Department of Education includes magnet and gifted/talented program schools in its identification process, it excludes these schools from the published list of beating-the-odds schools, even though they may be identified as beating the odds schools by one or both methods. Following Michigan's approach, the study team used data from all schools, including magnet schools. The study team also computed agreement rates with all schools, without excluding magnet schools from the identified lists. All schools used in the identification process were included in the computation of agreement rates in order to examine the variation in the identification results as originally generated by each method.

Comparing identification results across alternative specifications

To compare school identification results across different specifications, the identification results of the baseline model were compared with the identification results of each alternative model (models B–E). The results based on the school year 2010/11 data are presented in the next section (see appendix B for results for other years).

Comparing identification results across years

In comparing the school identification results across years, the study team documented differences in schools identified as beating the odds due to use of data from different school years. For each method the results based on the alternative model were produced for each of the past four school years (2007/08–2010/11) and compared across adjacent years. In addition, the frequency with which each school was identified as beating the odds over the four years was examined.

Table 2. Baseline and alternative models used with the prediction and comparison methods

Data and sample specification choice	Model A Baseline model (2010/11)	Model B Alternative performance measure model (2010/11)	Model C Alternative school characteristics model (2010/11)	Model D Alternative school configuration model (2010/11)	Model E Alternative performance measure and school characteristics and configuration model (2007/08 2010/11)
Performance measure					
Top-to-Bottom ranking percentile (available only for 2010/11)	✔		✔	✔	
Alternative composite index based on math/reading scale scores		✔			✔
School characteristics					
Baseline (different across methods)[a]	✔	✔		✔	
Alternative (comparable across methods)[b]			✔		✔
School sample configurations					
All levels pooled together	✔	✔	✔		
By school level[c]				✔	✔

a. The prediction method includes the following school characteristics as the baseline: percent English language learner students, percent eligible for free or reduced-price lunch, percent racial/ethnic minority, and percent with disabilities. In addition to those indicators, the comparison method includes the following school characteristics as the baseline: grades served, locale, total enrollment, special education center status, and state foundation allowance status.

b. Alternative school characteristics were selected based on the statistical significance of the regression coefficients in ordinary least squares estimation of school-level performance measures.

c. Identify beating-the-odds schools separately for elementary, middle, and high school grades.

Source: Authors' analysis based on Michigan Department of Education data.

The adjusted agreement rate shows the variation in beating-the-odds school identification results

To measure the differences and similarities between two identified school lists, the study team defined an adjusted agreement rate, computed as a ratio of the number of schools that appear in both lists to the average number of schools on a list. This agreement rate provides a measure of the extent to which a list includes commonly identified schools, adjusting for the size variation in the compared lists. This measure captures the share of commonly identified schools per list and highlights the degree of variation across alternative lists. For example, if there are 30 beating-the-odds schools on one list and 50 beating-the-odds schools on another list, and 10 schools appearing on both lists, the agreement rate is computed as 10/[(50+30)/2] = 0.25, or 25 percent. The agreement rate ranges from 0 to 100 percent. It attains the maximum possible value of 100 percent only if the two lists have exactly the same number of schools and exactly the same schools.

What the study found

This study demonstrates that different analytic decisions can lead to multiple sets of results. The findings show that variation in identifying beating-the-odds schools is a likely outcome when using different statistical methods or technical specifications regarding performance measures and school characteristics or applying different years of data.

Agreement rates in beating-the-odds schools between the two methods (prediction and comparison) were low, regardless of whether baseline or alternative specifications were used. Given a statistical method, the lists of beating-the-odds schools also varied depending on specifications applied. Variation in the identification results was particularly large when the performance measures were altered. For the prediction method, switching from the baseline performance measure to the alternative performance measure cut the number of schools identified in half. The list of schools identified as beating the odds would change over time to a considerable extent: overall, fewer than half the schools were identified more than once in the four-year period by either method.

This section highlights findings from between-method, within-year comparisons; within-method, within-year comparisons; and within-method, between-year comparisons of the schools identified for school years 2007/08–2010/11. Each beating-the-odds school list generated for these comparisons is specific to a combination of method, model, and year of data. The comparisons that most directly address the research questions are considered here. For the first two research questions that examine within-year comparisons, results presented below are limited to school year 2010/11 because that year's comparisons better align with Michigan's current focus on identification through the Top-to-Bottom ranking. Results from all comparisons are presented in appendix B.

Between-method comparisons of the prediction method and the comparison method using 2010/11 data yielded agreement rates in beating-the-odds schools identified of less than 50 percent for both baseline and alternative specifications

The prediction method and comparison method identified different sets of beating-the-odds schools

Between-method comparisons of the prediction method and the comparison method (research question 1) using 2010/11 data yielded agreement rates in beating-the-odds schools identified of less than 50 percent for both baseline and alternative specifications (table 3).

The baseline specifications (model A) mimic Michigan's choices in 2010/11, which included different school characteristics across the two methods (see box 1). Under the baseline specifications, the comparison method identified fewer than half the number of schools that the prediction method identified (28 versus 75), with 39 percent agreement.

The baseline analysis sample used for the two methods was not identical because of differences in missing data patterns associated with the different measures needed for each method. Therefore, the difference in the schools identified as beating the odds by the two methods might be caused by their differing analytic samples.

Table 3. Beating-the-odds school identification results varied by prediction and comparison methods, 2010/11

Model	Number of schools identified by			Agreement rate between methods (percent)
	Prediction method	Comparison method	Both methods	
A	75	28	20	39
E	71	35	17	32

Note: The total number of schools in the school year 2010/11 study sample is 3,563. For model A, the number of schools included in the identification is 3,325 for the prediction method and 3,279 for the comparison method. For model E, the number of schools included in the identification is 3,300 for both methods.

Source: Authors' analysis based on Michigan Department of Education data.

The alternative specifications (model E) used the same performance measures and an equivalent set of school characteristics. Under model E, the comparison method identified approximately half the number of schools that the prediction method identified (35 versus 71), with 32 percent agreement (see table 3). Because the same data are required by both methods to create outcome and demographic measures used in school identification, the analytic samples used for each method are identical.

The relatively low agreement rates in beating-the-odds schools identified between methods, using baseline or alternative specifications, suggests that the use of different statistical methods—prediction versus comparison—rather than the specification choices are likely driving the different school identification results.[3]

Some technical specification options influenced the list of schools identified as beating the odds

Within-method comparisons explored how the schools identified by a given method vary when alternative specifications are used for school performance measures, school characteristics, and school sample configuration (research question 2).

Switching from the baseline specification for the performance measure to the alternative performance measure cut the number of schools identified by half for the prediction method and more than doubled the number of schools identified for the comparison method

Within-method, within-year comparisons using 2010/11 data showed how the lists of schools identified as beating the odds differed between the baseline and alternative specifications. The largest differences occurred when the performance measures were altered (table 4). For the prediction method, switching from the baseline specification for the performance measure (model A Top-to-Bottom ranking) to the alternative performance measure (model B composite performance index) cut the number of schools identified by half. The agreement rate in schools identified was 11 percent. For the comparison method, switching from the baseline to the alternative performance measure more than doubled the number of schools identified, resulting in an agreement rate of 18 percent.

Table 4. Different performance measures caused the largest difference in beating-the-odds school identification results among the three specifications, 2010/2011

Method and performance measure	Model A Baseline model (2010/11)	Model B Alternative performance measure model (2010/11)	Model C Alternative school characteristics model (2010/11)	Model D Alternative school configuration model (2010/11)
Prediction method				
Number of schools identified as beating the odds	75	37	71	75
Number of schools overlapped with baseline model (model A)	na	6	55	75
Agreement rate with baseline model A (percent)	na	11	75	100
Comparison method				
Number of schools identified as beating the odds	28	70	30	30
Number of schools overlapped with baseline model (model A)	na	9	12	26
Agreement rate with baseline model A (percent)	na	18	41	90

na is not applicable.

Note: The total number of schools in the 2010/11 study sample is 3,563. For the prediction method, the number of schools included in the identification is 2,888 for model A, 3,300 for model B, 2,887 for model C, and 2,888 for model D. For the comparison method, the number of schools included in the identification is 2,791 for model A, 3,231 for model B, 2,887 for model C, and 2,888 for model D.

Source: Authors' analysis based on Michigan Department of Education data.

As mentioned earlier, the analysis sample size varies by model. Models A and B have considerable differences in sample size (2,888 versus 3,300 for the prediction method and 2,791 versus 3,231 for the comparison method). This is largely because of the unavailability of the Top-to-Bottom ranking data used in model A. That is, there were more missing outcomes when using Top-to-Bottom ranking data than when using alternative performance measures. The variation in the results between models A and B might be attributable to the difference in the pool of schools included. However, it is unlikely that the observed variation in the results is due solely to differences in the analysis sample. Additional data in appendix B show that even when comparisons are based on more similar performance measures and with much smaller sample size differences, the agreement rates were 53 percent for the prediction method and 28 percent or greater for the comparison method in 2010/11 (see table B12, model A' versus B').

Different choices of school characteristics lead to different beating-the-odds designations. Lower agreement for the comparison method suggests that school characteristic choices might be especially important for an approach that seeks to identify demographically similar schools for comparison

The choice of school performance measures greatly influenced the beating-the-odds school identification results. The findings in table 4 indicate the importance of determining which performance measure should be selected to identify beating-the-odds schools. The baseline and alternative performance measures selected for this study derive from the same overall state data, but they measure performance in different ways. The Top-to-Bottom ranking assigns an ordinal rank to schools that might not accurately measure performance differences between schools (for example, a school ranked 5th might have actual performance levels closer to the school ranked 20th than the school ranked 1st does). The alternative composite performance measure has an interval scale, reflecting quantifiable differences among schools with different composite sores. Changing performance indicators could change a given school's likelihood of meeting beating-the-odds criteria. Furthermore, performance measures of policy interest may not be uniformly available for all schools intended to be included in beating-the-odds school identification. In such cases the selection of a particular performance measure could further affect beating-the-odds school results by reducing (or expanding) the pool of schools available for consideration.

The selection of school characteristics also influenced beating-the-odds school identification results. The baseline and alternative specifications for school characteristics also produced variation in schools identified as beating the odds, though to a lesser degree than the performance measures, for the models presented in table 4. Lists of schools generated using the baseline school characteristics used by Michigan (model A) and the set of common school characteristics constructed by the study team (model C) contained a similar number of schools with a 75 percent agreement rate using the prediction method and a 41 percent agreement rate using the comparison method. These results indicate that different choices of school characteristics lead to different beating-the-odds designations. The lower level of agreement for the comparison method suggests that school characteristic choices might be especially important for an approach that seeks to identify demographically similar schools for comparison.

School sample configuration had little influence on beating-the-odds identification results. Comparing a beating-the-odds school list based on data in which schools were pooled regardless of grade-level configuration (model A) and the alternative list of schools based on data that compared schools with similar grade-level designations (model D) revealed minor differences in schools identified as beating the odds. The prediction method identified the same schools (100 percent agreement rate between results for all schools and schools by grade-level designations), and the comparison method identified

a similar set of schools (90 percent agreement rate). These findings suggest that decisions about comparing schools within or across grade levels are not as consequential for identifying beating-the-odds schools as the other specifications examined with Michigan data.

Less than half of the schools were identified as beating the odds in more than one year

Between-years comparisons explored how the schools identified as beating the odds by a given method varied from year to year (research question 3). Based on within-method, between-year comparisons, the agreement rates for schools identified in consecutive years averaged 48 percent for the prediction method and 49 percent for the comparison method (table 5). These agreement rates derived from comparable specifications applied to each method (model E) across all years, suggesting that year-to-year variation in the schools identified as beating the odds reflects changes in the underlying school performance and characteristic data or statistical noise rather than changes in analytic decisions.

Based on within-method, between-year comparisons, the agreement rates for schools identified in consecutive years averaged 48 percent for the prediction method and 49 percent for the comparison method

The year-to-year variation may be attributable in part to changes in the school sample over time and to changes in patterns of missing data. While the target sample, defined as public K–12 schools in Michigan, is the same each year, the size of the analytic sample used in year-to-year comparisons varied, ranging from 3,300 in 2010/11 to 3,490 in 2007/08. This sample size variation is likely to reflect underlying changes, such as school closures and new school openings. The analytic sample size variation may also reflect changes in missing data within schools. The between-years variation illustrated in table 5 (and table 6) reflects these differences in the analytic samples as well as the year-to-year variation in school performance and characteristics.

An examination of how frequently individual schools were identified as beating the odds found that, among schools identified at least once, only 9 percent were identified in all four years using the prediction method (16 out of 186) or the comparison method (15 out of 175). Overall, fewer than half of the schools were identified more than once in the four-year period by either method (see table 6). This indicates that relative school performance

Table 5. Both methods produced less than 50 percent agreement in schools identified as beating the odds over two years

Compared years		Number of schools identified			Agreement rate between periods (percent)
Period 1	Period 2	Period 1	Period 2	Both periods	
Prediction method					
2007/08	2008/09	86	85	36	42
2008/09	2009/10	85	82	43	52
2009/10	2010/11	82	71	38	50
Average of adjacent-year comparisons					48
Comparison method					
2007/08	2008/09	65	78	34	48
2008/09	2009/10	78	71	34	46
2009/10	2010/11	71	80	40	53
Average of adjacent-year comparisons					49

Note: Identifications are based on the alternative comparable specification (model E). For both the prediction and comparison methods, the number of schools included in the identification are 3,490 in 2007/08, 3,469 in 2008/09, 3,398 in 2009/10, and 3,300 in 2010/11.

Source: Authors' analysis based on Michigan Department of Education data.

Table 6. Fewer than half the schools were identified more than once over four years under either method

Number of years identified as beating the odds school	Prediction method	Comparison method
Number of schools identified in at least one of four recent school years (2007/08– 2010/11)	186	175
All four years	16	15
Three years	23	16
Two years	44	42
One year	103	102

Note: Identifications are based on the alternative comparable specification (model E).

Source: Authors' analysis based on Michigan Department of Education data.

levels, as measured in this study, change across years and contribute to turnover and inconsistency in schools on the beating-the-odds lists.

Implications of the study for identifying beating-the-odds schools

In response to the Beating the Odds Research Alliance's concerns that schools identified as beating the odds by the Michigan Department of Education varied more widely than expected, the study examined the extent of variation that can be expected when statistical methods, technical specifications, and time periods are changed. The study does not recommend one method or one set of technical specifications over another, but rather is designed to inform the process of developing and evaluating a technical approach to identifying beating-the-odds schools.

Variation in identification results for schools in Michigan is a likely outcome when different statistical methods or different technical specifications of performance measures, school characteristics, and school sample configuration are used, as well as when data for different years are applied

The findings show that variation in identification results for schools in Michigan is a likely outcome when different statistical methods or different technical specifications of performance measures, school characteristics, and school sample configuration are used, as well as when data for different years are applied. This study thus demonstrates that different analytic decisions can lead to multiple sets of results.

Recognizing a beating-the-odds school requires careful consideration by policymakers because of the many possible approaches to identifying such schools

The choices of statistical methods and technical specifications used to identify beating-the-odds schools reflect not only technical preferences but also specific definitions of "beating the odds." Although identification results are a product of a statistical process, they ultimately reflect policy decisions, and involving policy-minded stakeholders as well as technical staff is critical to developing a process that leads to meaningful identification of beating-the-odds schools.

The findings offer practical considerations for those developing or modifying a process for identifying beating-the-odds schools

The study suggests that policymakers or researchers engaged in developing or modifying a process for identifying beating-the-odds schools consider the following findings:

- Using multiple criteria to identify beating-the-odds schools may be beneficial, recognizing that any single method will have limitations.

- In developing an identification process, a range of definitions, methods, and technical specifications should be explored because the results could be highly sensitive to those choices. For example, identifying beating-the-odds schools by directly modeling a school effect (similar to value-added modeling) may be an alternative to using the prediction and comparison methods discussed in this study.
- The choice of a school performance measure is an important aspect of identifying beating-the-odds schools. Because there are many ways to define and measure school performance, the choice of a performance measure should reflect policy goals and consideration of the measure's validity, stakeholder support, and availability across years.
- Depending on policy priorities and preferences, an identification process may adjust statistical thresholds to determine whether a school is beating the odds or may modify measures of school characteristics to use in identification.
- Identification results based on a single year's performance data may be highly variable across years; this variability may be caused partly by factors beyond the schools' control. To reduce such variability, an identification process might incorporate performance data over multiple years. Depending on policy priorities and preferences, users might adjust statistical thresholds to determine whether a school is consistently beating the odds, modify measures of school characteristics to reflect information over multiple years, or consider schools' pattern of improvement over time.

Using multiple criteria to identify beating-the-odds schools may be beneficial, recognizing that any single method will have limitations

Limitations of the study

The study team notes five important caveats.

First, the analyses guided by the research questions are limited to documenting the observed range of variation when different statistical methods and technical specifications are used. The study does not determine the validity of either statistical method used by the Michigan Department of Education, nor does it establish the appropriateness of particular data and sample assumptions. The study is restricted to providing information on patterns in lists of schools identified as beating the odds arising from the two methods used and, for each method, from a set of alternative specifications.

Second, the study explored only a limited number of alternative specifications of the methods, focusing on demonstrating the resulting variation in the lists of schools identified as beating the odds. There was no comprehensive review of potential factors contributing to the variation in identification results. Future studies might explore potential sources of instability, which have not been explicitly investigated in this study.

Third, the study used an agreement rate as the primary measure to assess the variation across lists of schools identified as beating the odds, focusing on the observed overlaps of identified schools. Future studies might examine additional ways to assess the variation across identification approaches. For example, one might explore the extent to which schools identified as beating the odds by only one approach are close to being identified as beating the odds by the other approach. Such investigation could provide additional insight into the extent of the variation in the results over different identification approaches.

Fourth, when the study made pairwise comparisons of the results, it simply compared them as they would have been generated under two separate sets of choices regarding methods,

specifications, and year of the data. Differences in school samples available for analysis are among the consequences of these choices, due to differences in missing data patterns associated with particular choices. These analytic sample differences may have contributed to the variation across the lists. The results based on different analysis samples were presented and compared because they realistically reflect how states are likely to identify beating-the-odds schools. However, the sample differences complicate the interpretation of the observed relationships between identification results and technical choices made. The study was not designed to identify which technical choice causes the variation in schools identified as beating the odds; therefore, further investigation and conclusions as to the causes of the variation is left to future studies.

Finally, findings presented in this report are based on a limited number of models selected to illustrate the potential extent of the variation in the identification of beating-the-odds schools.[4] These findings do not focus on particular identification results, but highlight the sensitivity of the results to the choice of the method and model selections. Readers are cautioned against viewing the particular identification results reported here as conclusive. The identification results based on additional models (see appendix B) show further variation, demonstrating that there are as many different identification results as there are models.

Appendix A. Literature review

A limited number of studies have developed methods for identifying beating-the-odds schools or districts, typically with a goal similar to Michigan's of identifying practices and policies that distinguish these schools and that might be replicable elsewhere. The methods of identifying beating-the-odds schools, as well as their level of sophistication, vary across the studies. Collectively, the studies highlight different decisions related to identifying schools, including defining the population of schools included for consideration, the demographic predictors or controls and performance criteria, the period of performance to be analyzed, and the choice of outcome measures.

The population of schools considered as beating the odds

An initial determination to be made in approaching identification of beating-the-odds schools concerns the population of schools to be included for consideration. Similar to the Michigan Department of Education's approach, a 2007 California study (Pérez et al., 2007) defined beating-the-odds schools as those performing statistically higher by a significant margin than expected given the population they serve. Accordingly, beating-the-odds schools were identified from among all public schools statewide (excluding charter and magnet schools to maintain comparability of schools).

However, in other studies beating-the-odds schools are defined specifically as high-poverty, high-performing schools. Reeves (2004, p. 186) employed the "90–90–90" criteria to identify beating-the-odds schools: "More than 90 percent of the students are eligible for free and reduced-price lunch. More than 90 percent of the students are from ethnic minorities.… More than 90 percent of the students met or achieved high academic standards." Similar criteria were used by Kearney, Herrington, & Aguilar (2012) in a Texas study. Studies by Mid-Continent Research for Education and Learning (for example, Apthorp et al., 2005) limited identification to schools with at least 50 percent of students eligible for free or reduced-price lunch. In Arizona, following state policy priorities, beating-the-odds schools were identified from among schools in which at least 50 percent of the students were eligible for free or reduced-price lunch and at least 50 percent were Latino (Waits et al., 2006). A Delaware study included middle schools only (Grusenmeyer et al., 2010).

Demographic predictors or controls and performance criteria for beating-the-odds schools

Several of these studies used one of the approaches adopted by Michigan. Similar to Michigan's prediction method, some studies identified beating-the-odds schools as those that outperformed their regression-predicted score, with predictions based on rates of poverty, proportions of English language learner students, proportion of students with disabilities, parent education levels, and other demographic characteristics (Apthorp et al., 2005; Pérez et al., 2007). In an approach comparable with Michigan's comparison method, Delaware researchers formed "clusters" of comparable schools and identified those that outperformed comparison schools (Grusenmeyer et al., 2010). School clusters were based on schoolwide percentages of White students and students from low-income households.

Attainment of annual performance growth targets was the key criterion for Cudeiro, Palumbo, and Nelsen (2005). This study identified six beating-the-odds schools in Southern California, all with high percentages of students eligible for free or reduced-price lunch,

racial/ethnic minority students, and English language learner students and all having met a 5 percent annual growth rate between 2000 and 2005 in the school's California Academic Performance Index (a weighted composite of all state-administered test scores), both schoolwide and for identified demographic groups.

Approaches to determining performance "cutoffs" for beating-the-odds schools varied across the studies, reflecting different levels of stringency and precision. Apthorp et al. (2005) applied a cutpoint for beating-the-odds schools of 0.75 standard deviation above or below the predicted score, while Pérez et al. (2007) set the cutpoint at 0.75 standard deviation above the mean residual, both for all students and for relevant subgroups (students eligible for free or reduced-price lunch, African-American students, Hispanic students, and English language learner students).

In the Delaware study (Grusenmeyer et al., 2010), beating-the-odds schools within each socioeconomic cluster were those with most or all of the targeted test scores (math and reading scores in the state testing program) higher than the cluster averages. In identifying high-poverty beating-the-odds elementary schools, the Education Priority Panel in New York City used "above city average" on a city-administered language arts test as a criterion (Connell, 1999). For the "90–90–90" approach, Kearney et al. (2012) and Reeves (2004) set the cutpoint as a 90 percent pass rate on state-mandated exams in math and English language arts.

School configuration as a consideration in identifying beating-the-odds schools

Although some studies did not consider the grade configuration of a school when determining beating-the-odds status, a few suggest that there could be important reasons to stratify identification by school configuration. Most notably, these studies limited their research to specific grade levels. For example, researchers of one New York City study focused on high schools, tracking how the schools prepared originally low-performing grade 9 students for college success (Ascher & Maguire, 2007). Researchers of another study of New York City schools focused on high-poverty, high-achieving elementary schools (Connell, 1999). Apthorp et al. (2005) used demographic information and school performance data to identify only elementary schools that were beating the odds. Pérez et al. (2007) studied beating-the-odds schools of all levels in California but varied the requirements for identification based on whether the school was an elementary, middle, or high school.

Time period analyzed in identification of beating-the-odds schools

A number of the studies reviewed identified beating-the-odds schools based on multiple years of performance. Delaware (Grusenmeyer et al., 2010), for example, required that beating-the-odds schools demonstrate consistent high performance for three years, while an Arizona study team (Waits et al., 2006) examined patterns of performance during an eight-year period. In it, identified schools were either "steady performers" (that is, they consistently outperformed their expected levels and performed above the statewide average for eight years) or "steady climbers" (that is, they showed a gain of at least 9.5 points on relevant measures while also avoiding any declines of more than 10 points during the eight-year period). In the California study, Pérez et al. (2007) used four years of test score data for elementary and middle schools and three years of test score data for high schools.

Several researchers noted that reliance on a single year of data is more likely to result in beating-the-odds identification based on factors other than school practices. In a study of schools that beat the odds in graduation rates, Socias, Dunn, Parrish, Muraki, and Woods (2007) describe the potential problem of year-specific effects; if a school experiences a particularly difficult year, for example, because of a large number of retirements, the school's performance may be misrepresented by a single year of data in the absence of historical data to smooth out the shock. Harris (2007) describes the regression to the mean caused by statistical noise (random error) that may result in misidentification of high-performing schools if multiple years, grades, and test scores are not analyzed. Because of statistical noise in standardized test scores, schools that appear to be high performing in a single year may in fact be performing at average or lower levels. In a critique of studies by The Education Trust (2005) and Carter (2000), Harris argues for use of multiple years of data (as well as multiple tests and grade levels) in determining "high flying"—high-performing, high-poverty—schools. Harris found that 93 percent of schools identified as high flying in The Education Trust approach, based on performance in a single year on a single subject, were not considered high flying when the identification standard was raised to require high performance across at least two years, two tests, and two grade levels.

Selecting outcome measures for identifying beating-the-odds schools

In several studies, state choices of outcome measures to be analyzed as a basis for identifying beating-the-odds schools were narrowed to follow state policy priorities. Arizona limited outcome measures to those considered "best reflections of critical junctures of learning": grade 3 reading and grade 8 math scores (Waits et al., 2006). Delaware used math and reading scores (on Delaware tests administered statewide) for grades 6–8. As indicated by Socias et al. (2007), cross-measure correlations may be relatively weak, implying wide variations in beating-the-odds school identification results depending on the measures used. Decisions about the measures are therefore highly consequential and inevitably reflect subjective preferences.

Additional considerations noted in the aforementioned studies included whether state assessments were administered consistently across schools, whether tests were vertically equated (if scores were comparable across grade levels), and whether the percentage of students tested each year varied considerably within schools. Pérez et al. (2007) noted that at the high school level, no single math test could be compared across schools at common grade levels because math tests varied depending on the courses in which students were enrolled; the California High School Exit Exam results for grade 10 students were therefore used at the high school level, while state-administered math and English language arts tests were used at the elementary and middle school levels.

Socias et al. (2007) also noted caution regarding use of measures that are sensitive to "cohort shock"—extreme fluctuations in performance that result from atypical cohorts of students. To analyze the stability of various measures over time, Socias et al. examined cross-year correlation (correlation of each measure with its value the previous year) of potential dropout measures. They recommended using the more stable measures to identify schools that are beating the odds through dropout prevention practices.

Screening or filtering schools for final identification as beating-the-odds schools

Beyond higher-than-expected achievement on standardized tests, some procedures for identifying beating-the-odds schools have required that schools meet other baselines such as minimum attendance or graduation rates or demonstration of inclusive enrollment. Harris (2007) points out that schools may be misidentified as beating the odds if they have restrictive admission policies. In the New York study, Connell (1999) screened out schools that recruited students through gifted and talented programs or that excluded low-performing students through high levels of special education referrals. Socias et al. (2007) determined through initial interviews that some schools identified through statistical analysis as beating the odds were "false positives"; these schools transferred problem students into alternative schools rather than working with them.

Identification of beating-the-odds schools based on specific practice

A more focused approach to beating the odds has been used by several researchers with an interest in effective approaches to reading instruction. Rather than relying purely on demographics and outcomes criteria, these researchers identified schools for study based on known reading initiatives or teaching efforts at the schools, as well as on outcomes. This approach provided an opportunity for in-depth research on reading instruction. To select schools for the intensive field study, Langer (2000) asked experts and practitioners for recommendations of schools in four states in which English teachers had been known to include in their professional duties the exertion of effort in improving reading achievement and in which attendance, enthusiasm for learning, and student achievement had improved. Langer identified beating-the-odds schools (schools that outperformed demographically similar schools) within those schools that were nominated, although all recommended schools were included in the field study. Taylor, Pearson, Clark, and Walpole (1999) identified high-poverty elementary schools that had recently implemented a program to increase reading achievement and that were also known for high achievement. Based on analysis of gain scores, a subset of those schools originally identified were categorized as beating-the-odds schools. To identify practices contributing to high reading scores, Taylor et al. compared the beating-the-odds and non-beating-the-odds schools within the sample through teacher and principal surveys and interviews.

Summary of key considerations for identification methods

This small body of literature on beating-the-odds school identification methods indicates that these methods reflect subjective decisions, policy priorities, and specific policy or research objectives for the beating-the-odds school identification. Given the goal of learning about school practices that lead to higher-than-expected performance, the literature underlines several considerations in designing methods to identify beating-the-odds schools that may have produced positive outcomes through their own school policies and practices. The inclusion of multiple years of performance data, measures that are comparable across schools and years, and schools that have been screened for "selective" enrollment is among these key considerations.

Appendix B. Technical details on methods and additional results

This appendix gives details on the data and analyses presented in the report. It also provides results from the identification of beating-the-odds schools based on additional specifications and years.

Data files

The analysis used school and student assessment and demographic records from kindergarten–grade 12 public schools in Michigan for the school years 2007/08, 2008/09, 2009/10, and 2010/11. The number of schools included in analyses by year is shown in table B1.

The data were made available by the Michigan Department of Education and supplemented with data from the National Center for Education Statistics Common Core of Data (U.S. Department of Education, 2009, 2010, 2011, 2012). The unit of analysis for the study was the school; student data were used to create school variables.

School-level records included:

- Michigan Department of Education-provided data
  - Special education center status.
  - Grades served by school (for example, K–5, 9–12).
  - Total enrollment for each school, by grade.
  - Number of students eligible for free or reduced-price lunch.
  - Number of students with disabilities.
  - Number of English language learner students.
  - Number of students by race/ethnicity.
  - Number of female students.
  - Four-year cohort graduation rate.
  - Four-year cohort dropout rate.

Table B1. Total schools in Michigan and number of schools included in analyses, by grades served and year

School group and grades served	2007/08	2008/09	2009/10	2010/11
All schools				
Total number of schools serving K–12	3,733	3,710	3,649	3,563
Number of schools serving grades K–5	2,253	2,234	2,188	2,136
Number of schools serving grades 6–8	1,511	1,503	1,472	1,417
Number of schools serving grades 9–12	1,149	1,159	1,159	1,196
Number of magnet schools	445	478	474	463
Schools included in analysis[a]				
Total number of schools serving K–12	3,547	3,508	3,441	3,436
Number of schools serving grades K–5	1,289	1,293	1,252	1,285
Number of schools serving grades 6–8	473	469	448	402
Number of schools serving grades 9–12	743	721	730	730
Number of magnet schools	431	464	462	457

a. Schools included in at least one identification model for either method.

Source: Authors' analysis based on Michigan Department of Education data.

    - School year 2010/11 Top-to-Bottom percentile ranking.
    - Michigan state foundation allowance.
- Common Core of Data
    - Type of school (regular, magnet, gifted, or other special program school).
    - National Center for Education Statistics geographic category.

Student records included:

- High school graduation indicator.
- High school dropout indicator.
- Economically disadvantaged (eligible for free or reduced-price lunch) indicator.
- Primary disability.
- Grade level.
- English language learner indicator.
- Race/ethnicity.
- Gender.
- Michigan Educational Assessment Program scale scores (grades 3–8), by content area (math, reading, science, social studies, and writing).[5]
- Michigan Educational Assessment Program proficiency level (grades 3–8), by content area.
- Michigan Merit Examination scale score (grade 11), by content area.
- Michigan Merit Examination proficiency level (grade 11), by content area.
- MI-Access scored points and scale score (students with disabilities), by content area.
- MI-Access proficiency level, by content area.
- ACT math scale score.
- ACT reading scale score.
- ACT English scale score.
- ACT science scale score.

Student data were aggregated to create the school data. For example, student English language learner status was aggregated up to the school level to create a percent English language learner student measure. The same sample was used for creating both the control and outcome measures. The sample used to create these measures varied because data were missing for some measures. Thus, the students whose assessment results are used to construct a performance indicator for a given school may not be exactly the same as the students whose demographic records are used to compute the school-level demographic indicators.

The summary descriptions of select school performance indicators as well as school demographic indicators for all K–12 public schools in Michigan are shown in table B2.

Study and analysis samples

All K–12 public schools in Michigan were included in the study. The beating-the-odds school identifications were conducted separately for each of the four study years (2007/08–2010/11). Where student records are used to create the school-level variables, all students for whom schools could be identified and records were available were included. The number of schools in the study sample by configuration (including total number of schools and number of schools by grade levels served), school characteristics, and average school assessment scores are shown in table B2.

Table B2. School configuration, characteristics, and average assessment scores for all K–12 schools in Michigan, 2007/08–2010/11

School configuration		2007/08			2008/09			2009/10			2010/11		
Total number of schools serving K–12		3,733			3,710			3,649			3,563		
Number of schools serving grades K–5		2,253			2,234			2,188			2,136		
Number of schools serving grades 6–8		1,511			1,503			1,472			1,417		
Number of schools serving grades 9–12		1,149			1,159			1,159			1,196		
Number of magnet schools		445			478			474			463		
School characteristics		**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**
Total enrollment[a]		3,733	443	344	3,710	437	337	3,649	442	338	3,563	439	337
Female ratio		3,733	0.47	0.09	3,710	0.47	0.09	3,649	0.47	0.09	3,563	0.47	0.10
Free or reduced-price lunch ratio		3,733	0.46	1.60	3,710	0.49	0.40	3,649	0.51	0.25	3,563	0.50	0.25
English language learner ratio		3,733	0.03	0.09	3,710	0.04	0.13	3,649	0.04	0.20	3,563	0.03	0.09
Students with disabilities ratio		3,733	0.13	0.13	3,710	0.14	0.13	3,649	0.13	0.13	3,563	0.13	0.14
Minority ratio[b]		3,733	0.09	0.12	3,710	0.10	0.12	3,649	0.10	0.12	3,563	0.11	0.13
Cohort graduation rate[c]		880	0.72	0.29	874	0.71	0.29	880	0.71	0.29	908	0.71	0.29
Cohort dropout rate[c]		887	0.17	0.20	861	0.14	0.17	885	0.16	0.19	956	0.16	0.20
Average school assessment scores		**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**
Math	Grade 5 (MEAP)	1,797	518	17	1,791	521	18	1,727	523	17	1,671	523	17
	Grade 8 (MEAP)	1,017	810	16	1,018	814	15	1,005	812	15	1,042	812	14
	Grade 11 (MME)	956	1,083	19	966	1,083	21	968	1,081	24	1,053	1,082	27
Reading	Grade 5 (MEAP)	1,795	527	16	1,791	525	16	1,728	529	14	1,672	528	14
	Grade 8 (MEAP)	1,018	812	15	1,018	815	15	1,007	819	12	1,041	817	14
	Grade 11 (MME)	956	1,096	19	966	1,096	18	968	1,098	19	1,057	1,100	21
Science	Grade 5 (MEAP)	1,797	522	16	1,791	525	16	1,729	523	15	1,674	521	17
	Grade 8 (MEAP)	1,017	817	16	1,018	815	16	1,004	814	14	1,038	815	14
	Grade 11 (MME)	957	1,088	22	965	1,087	23	968	1,089	23	1,053	1,093	26
Social science	Grade 6 (MEAP)	1,190	610	14	1,167	611	13	1,142	611	13	1,150	609	11
	Grade 9 (MEAP)	903	910	15	907	912	15	908	911	15	968	910	13
	Grade 11 (MME)	962	1,115	15	968	1,117	16	973	1,115	14	1053	1116	16

SD is standard deviation. MEAP is Michigan Educational Assessment Program; MME is Michigan Merit Examination.

Note: Not all schools are included in the study sample. See table B1 for the size of the study sample and table B3 for the size of the analytic sample for each model.

a. Derived by taking the sum of enrolled students from all grades.

b. Includes American Indian, Asian, African-American, Native Hawaiian, Hispanic, and students with multiple races.

c. Calculated by the Michigan Department of Education by tracking students starting at their enrollment as grade 9 students, with a completion or dropout rate over four years.

Source: Authors' analysis based on Michigan Department of Education data.

Table B3. Analysis sample size for identification by model and method, 2007/08—2010/11

Model specifications	Model A	Model B	Model C	Model D	Model (C+D)
Outcome	Top to Bottom	Alternative measure	Top to Bottom	Top to Bottom	Top to Bottom
School characteristics	Michigan-selected	Michigan-selected	Alternative best fit	Michigan-selected	Alternative best fit
Sample configuration	Pooled	Pooled	Pooled	By school level	By school level
Prediction method					
2010/11	2,888	3,300	2,887	2,888	2,888
Comparison method					
2010/11	2,791	3,231	2,887	2,888	2,888
Model specifications	**Model A'**	**Model B'**	**Model C'**	**Model D'**	**Model E/E'**
Outcome	Michigan-defined measures	Alternative measure	Alternative measure	Alternative measure	Alternative measure
School characteristics	Michigan-selected	Michigan-selected	Alternative best fit	Michigan-selected	Alternative best fit
Sample configuration	Pooled	Pooled	Pooled	By school	By school
Prediction method					
2007/08	3,512	3,490	3,490	3,490	3,490
2008/09	3,489	3,470	3,469	3,470	3,469
2009/10	3,424	3,399	3,398	3,399	3,398
2010/11	3,325	3,300	3,300	3,300	3,300
Comparison method					
2007/08	3,462	3,424	3,490	3,424	3,490
2008/09	3,437	3,401	3,469	3,401	3,469
2009/10	3,377	3,335	3,398	3,335	3,398
2010/11	3,279	3,231	3,300	3,231	3,300

Source: Authors' analysis based on Michigan Department of Education data.

Although all K–12 public schools are included in the study, the actual sample used in each analysis of beating-the-odds school identification varies by method and model. This is because the data elements for outcome or school demographics required for identification vary by method and model and because missing data patterns for the outcome and school demographic data elements also vary by school. For example, a school that is missing the Top-to-Bottom ranking may not be included in an identification model using Top-to-Bottom ranking as the outcome measure, but it may be included in another model not requiring Top-to-Bottom ranking. The study team did not impute missing data. The analytic sample size by model and method for each school year are summarized in table B3.

One implication of the varying analytic samples across models and methods is that they could contribute to the observed variation in the school identification results. As shown in table B3, the sample size varied across models, especially between those based on using Top-to-Bottom ranking as the outcome measures and those based on Michigan Department of Education or study team–defined composite performance measures in 2010/11. This is because a large number of schools were missing Top-to-Bottom ranking information (Top-to-Bottom ranking was not available prior to 2010/11).

Performance measures

The study team employed the following performance measures:

- *Top-to-Bottom percentile ranking* (used as the baseline measure for 2010/11). The Top-to-Bottom ranking is based on a performance index developed by the

Michigan Department of Education that takes into account the level and changes in school academic achievement, graduation rates, and within-school achievement gaps (see box 2 in the main report for more details). Top-to-Bottom ranking is not available prior to 2010/11.

- *Michigan Department of Education–developed performance measure for the prediction method* (referred to as the "initial" or "pre-2010/11" measure and used as the baseline performance measure for school years 2007/08–2009/10). Prior to 2010/11, Michigan used a composite performance measure specifically developed to be used with the prediction method for identifying beating-the-odds schools. This measure is constructed from student standardized assessment scores based on Michigan state tests: the Michigan Educational Assessment Program for elementary and middle school students, the Michigan Merit Examination for high school students, and the MI-Access for disabled students. Details on this measure are provided in the "Constructing school performance measures" section in this appendix.
- *Michigan-developed performance measure for the comparison method* (also called the "initial" or "pre- 2010/11" measure, used as the baseline performance measure for school years 2007/08–2009/10). Prior to 2010/11, the department developed this performance measure to be used specifically with the comparison method. It is based on the following published indicators: percentage of students meeting ACT college-readiness benchmarks in high school; percentage proficient in math, reading, science, social studies, and writing on Michigan Merit Examination or MI-Access in high school; percentage proficient on the same content areas on Michigan Educational Assessment Program or MI-Access in elementary or middle school; and graduation rates and dropout rates in high school. Details of this measure are provided in the "Constructing school performance measures" section in this appendix.
- *Alternative performance measure.* The study team constructed a composite performance index from student standardized assessment scores based on Michigan state math and reading tests. This alternative performance measure is a modified version of the performance index developed previously by Michigan to be used with the prediction method (see above).

The results based on the Top-to-Bottom ranking and author-developed alternative performance measure are discussed in the main report.

Beating–the-odds identification steps

Prediction method. For the prediction method, the first step was to construct performance measures (ordinary least squares linear model) as dependent variables. The baseline measure (the Top-to-Bottom ranking for 2010/11 and the Michigan Department of Education–developed "initial" [pre-2010/11] measures) and the alternative performance measure (the school average of *z*-score scores across two core subjects—math and reading) were prepared to be used as a dependent variables in the prediction model.

To construct an initial or alternative performance measure, student-level *z*-scores by content area were first computed based on the assessment data for all students in the state by grade for a given year; then the content area school mean was computed by taking the average of the student *z*-scores for each school for the given year; and, finally, the performance measure was calculated by taking the overall mean of the school average *z*-scores across all content areas and grades for each school for the given year.

The second step was to prepare a set of school demographic characteristic variables to be used as the covariates in the prediction model. For the baseline model (model A), these covariates included school demographic variables for the percentage of students eligible for free or reduced-price lunch, students with disabilities, and English language learner students. In addition to these variables, the alternative set of covariates included percent female, percent racial/ethnic minority, total enrollment, grades served, magnet school indicator, and locale indicators.

The third step was to regress, for each model, a school performance measure on a set of school demographic variables. The prediction model is expressed as:

$$\hat{Y}_j = \hat{\beta}_0 + \sum_m^M \hat{\beta}_m X_{mj} \quad \text{(B1)}$$

where Y represents a performance level of school j, X_m represents the m-th school demographic characteristic (m = 1 to M), and β represents a regression coefficient from the prediction model.

The fourth step was to calculate a predicted value for the performance measure for each school, based on the estimated prediction model. The prediction method identifies a beating-the-odds school when the actual performance exceeds the predicted performance by a certain margin. There is no one right answer to a question of how much the predicted value must exceed the actual value for a school to be considered a beating-the-odds school. The study team followed the Michigan Department of Education's identification criteria, which required that the predicted value exceed the actual value at least by two times the root mean square error from the prediction model. The cutoff point for the identification criteria corresponds roughly to the upper bound of the 95 percent confidence interval.

Comparison method. For the comparison method, the first step was to construct school performance measures to be used in comparing a school with demographically similar schools. The comparison method used the same sets of performance measures as the prediction method: the baseline measure (the Top-to-Bottom ranking for school year 2010/11 or the department of education–developed "initial" [pre-2010/11] measure) or the alternative performance measure (the school average of z-score scores across the core subjects).

The second step was to prepare a set of school demographic characteristic variables to be used to identify demographically similar schools. The variables used in the baseline model for the comparison method differ from those used in the prediction method and include the following (with weights used by Michigan in 2010/11 in parentheses: a variable weighted 5 is given five times the weight of one weighted 1): percentage of students eligible for free or reduced-price lunch (5), percentage of students with disabilities (2), percentage of English language learner students (3), percentage of racial/ethnic minority students (1), indicators for locale (1), total number of tested students (1),[6] indicators for school configuration (1), special education center status (10), and state foundation allowance amounts (1). The weights for each school characteristic were used in calculating Euclidean distance that is presented in the next paragraph and equation B2 (as w_d).The alternative set of characteristics used in the comparison method is the same as the alternative characteristics used by the prediction method. These alternative characteristics each had a weight of 1 in the calculations.[7]

The third step was to compute a weighted Euclidean distance measure between a given school and every other school in the sample. The distance (that is, the measure of how similar two schools are) was calculated as follows:

$$distance_{jk} = \sqrt{\frac{\sum_{d=1}^{N_d} w_d (z_{dj} - z_{dk})^2}{\sum_{i=1}^{N_d} w_d}} \quad \text{(B2)}$$

where $distance_{jk}$ is the distance between schools j and k in terms of demographics, N_d is the number of school characteristics which schools are compared, w_d is the weight placed on a characteristic d, z_{dj} is the z-score of school j on characteristic d, and z_{dk} is the z-score of school k on characteristic d.

The fourth step was to select a group of demographically similar schools as a comparison group. Based on the calculated distances, schools were ranked by how far they were from each given school. The shorter the distance between two schools, the more demographically similar they are. There is no one right criteria for choosing demographically similar schools for a given school or for deciding the extent by which the school's performance needs to exceed other similar schools' performance. As with the prediction method, the study team followed the Michigan approach, which selects 29 demographically similar comparison schools[8] for each school based on the ranking of their distance from this school.

If schools are very similar, they are likely to be in one another's comparison groups; however, the comparison groups are unique to each school and unlikely to be identical even for similar schools (unless schools are virtually identical on every demographic parameter). One school's identification as a beating-the-odds school does not preclude another school within its comparison group from also being identified as a beating-the-odds school because the inverse might not be true. That is, school B might be in school A's comparison group, but school A might not be in school B's comparison group.

The fifth and final step was to compare each school's performance with the performance of demographically similar schools identified in step four. If the school's performance measure is higher than all other similar schools and is statistically significantly higher than the comparison group mean by at least two times the comparison group standard deviation (at the $\alpha = 0.05$ level), it is identified as a beating-the-odds school. The cutoff point roughly corresponds to the upper bound of a 95 percent confidence interval.

Measures for the variation in beating-the-odds school identification results

To measure the differences and similarities between two beating-the-odds school lists, the study team computed an agreement rate, which is defined as a ratio of the number of schools that appear on both sets of the lists to the average number of schools across the two lists. The agreement rate R_{ij} between school i and school j can be expressed as follows[9]:

$$R_{ij} = \frac{n_{ij}}{(N_i + N_j)/2} \quad \text{(B3)}$$

where N_i is the number of schools on beating-the-odds list i, N_j is the number of schools on beating-the-odds list j, and n_{ij} is the number of schools included on both lists i and j.

This agreement rate provides a measure of the extent to which a list includes commonly identified schools, adjusting for the size variation in the compared lists. This measure captures the share of commonly identified schools per list and highlights the degree of variation across alternative lists. The agreement rate ranges from 0 to 100 percent. It attains the maximum possible value of 100 percent only if the two lists have exactly the same number of schools and exactly the same schools.

Baseline and alternative data and sample specifications applied to both methods

Key alternative data and sample specifications examined in the study are summarized in tables B4 and B5. They are the data equivalent of box 2 in the main report. Each set of specifications is referred to as a model.

Following Michigan's current approach for school year 2010/11, the study team first generated baseline beating-the-odds school lists based on both methods, using the Top-to-Bottom ranking as the performance measure and the department of education's set of school characteristics for each method (table B4, model A) and pooling the sample across school grade levels. Then, the study team generated beating-the-odds school lists by altering just one of the three specification items at a time—the performance measure (model B), the choice of school characteristics (model C), and the school sample configuration (model D), while keeping the other two specification items unchanged from the baseline model. Finally, the study generated beating-the-odds school lists by altering all three specification items at the same time (model E).

In addition, for each school year prior to 2010/11, the study conducted a similar investigation of the alternative specifications. Because the Top-to-Bottom ranking information was not available prior to 2010/11, the alternative composite performance index was applied when pre-2010/11 data were used. The key alternative data and sample specifications applied for data for school years 2007/08–2009/10 are summarized in table B5.

Constructing school performance measures

Beating-the-odds status is based on school performance measures. As noted earlier, this study applied four different types of school performance measures. The main report presents the identification results based on two of those measures: Top-to-Bottom ranking (percentile), which has been published by the Michigan Department of Education since 2010/11, and an author-defined composite academic performance index, computed as the average of *z*-scores across two core subjects (math and reading). The composite performance index was used with 2007/08–2010/11 data and is a modified version of the performance measures used by the Michigan Department of Education for the prediction method prior to school year 2010/11.

Top-to-Bottom ranking. As noted in the main report, since 2010/11, Michigan has used the Top-to-Bottom school percentile ranking published by the state as the primary performance measure in the beating-the-odds school identification. Top-to-Bottom takes into account the five tested areas (math, reading, science, social studies, and writing) of student assessment, as well as graduation and dropout rates and year-to-year achievement. Michigan uses the raw ranking, which ranges from 0 to 99, as the outcome variable in its beating-the-odds models. Prior to 2010/11 Michigan used performance measures constructed by the department separately for the prediction method and comparison method. That is, the initial performance measures used by Michigan are not directly comparable between methods.

Table B4. Baseline and alternative model specifications, 2010/11

Specification items	Model A Baseline model	Model B Alternative performance measure model	Model C Alternative school characteristics model	Model D Alternative sample model	Model E Alternative performance measure, school characteristics, and sample
Performance measure					
Top top-to-bottom ranking percentile	✔		✔	✔	
Alternative composite index based on standardized math and reading scale scores		✔			✔
School characteristic					
Original (different across methods)[a]	✔	✔		✔	
Alternative (comparable across methods)[b]			✔		✔
School sample configuration					
Grades pooled	✔	✔	✔		
Grade levels separated[c]				✔	✔

a. For the prediction method, the original set includes percent racial/ethnic minority for school year 2010/11. For the comparison method, the previous set excludes locale indicators for 2010/11.

b. Alternative school characteristics were selected based on the statistical significance of the regression coefficients in an ordinary least squares estimation of school-level performance measures.

c. Schools identified by elementary, middle, and high school grades.

Source: Authors' analysis based on Michigan Department of Education data.

Table B5. Prebaseline and alternative model specifications, 2007/08–2009/10

Specification items	Model A' Prebaseline (initial Michigan-developed measures)	Model B' Alternative performance measure model	Model C' Alternative school characteristics model	Model D' Alternative sample model	Model E' Alternative performance measure, school characteristics, and sample
Performance measure					
Initial outcome measures (Michigan Department of Education–created measures, different by methods)	✔				
Alternative composite index based on standardized math and reading scale scores		✔	✔	✔	✔
School characteristic					
Original (different across methods)[a]	✔	✔		✔	
Alternative (comparable across methods)[b]			✔		✔
School sample configuration					
Grades pooled	✔	✔	✔		
Grade levels separated[c]				✔	✔

a. For the prediction method, the original set does not include percent racial/ethnic minority for pre-2010/11. For the comparison method, the previous set includes locale indicators for pre-2010/11.

b. Alternative school characteristics are selected based on the statistical significance of the regression coefficients in an ordinary least squares estimation of school-level performance measures.

c. Schools identified by elementary, middle, and high school grades.

Source: Authors' analysis based on Michigan Department of Education data.

Initial (pre-2010/11) Michigan-defined performance measures. The initial (pre-2010/11) Michigan-defined performance measures were constructed based on a group of school assessment scores (scores from different subjects and different types of tests) by first creating a standardized measure on each test and then taking the simple mean across all standardized measures, without considering the number of students who took tests. These measures differ across the two methods in that each method used a different set of variables to create the composite outcome measures. The measure used for the prediction method standardized student scores on each test, while the one used for the comparison method standardized school percent proficiency on each test, school percent proficiency on each test from the previous testing year, plus cohort graduation and dropout rates for each school.

Prediction method. More specifically, the initial (pre-2010/11) measure for the prediction method was based on student standardized assessment scores on state tests, including the Michigan Educational Assessment Program for elementary and middle school students, the Michigan Merit Examination for high school students, and the MI-Access for students with disabilities. The content areas tested, grades tested, and the assessments used in constructing the pre-2010/11 prediction method performance measure are shown in table B6.

The composite performance measure was constructed as follows: first, computing student *z*-scores for each content area (math, reading, science, social studies, and writing)[10] of each assessment based on all student data in the state by grade; next, taking the average of the student *z*-scores in each content area for each school and creating a school content area performance index; and, finally, calculating the overall mean of the average *z*-scores across all content areas for each school. Specifically, for the prediction method, the performance measure (Y_j) for each school (j) was computed as follows:

$$Y_j = \frac{\sum_{k=1}^{N_{jk}} \left(\frac{\sum_{i=1}^{N_{ijk}} z_{ijk}}{N_{ijk}} \right)}{N_{jk}} \qquad \text{(B4)}$$

where k is an indicator representing each content area for which school j has data, N_{jk} is the number of content areas in which school j has data, i is an indicator representing each

Table B6. Assessments used in performance measures

Assessment	Content area	Grades tested
Michigan Educational Assessment Program and Michigan Merit Examination	Math	3–8, 11
	Reading	3–8, 11
	Science	5, 8
	Social studies	6, 9
	Writing	Prior to 2009/10: 3–8, 11 2009/10: None (because of field testing of new items) Post-2009/10: 4, 7
MI-Access participation and supported independence	Math	3–8, 11
	Reading	3–8, 11
MI-Access functional independence	Math	3–8, 11
	Reading	3–8, 11

Source: Authors' compilation.

student in school j, N_{ijk} is the number of students with scores on subject k in school j, and z_{jk} is the z-score of student i on content area k in school j.

Comparison method. The initial (pre-2010/11) measure for the comparison method was also a composite index, based on the following published indicators: percentage meeting ACT college-readiness benchmarks in high school; percentage proficient in math, reading, science, social studies, and writing on the Michigan Merit Examination or MI-Access in high school; percentage proficient on the same content areas on Michigan Educational Assessment Program or MI-Access in elementary or middle school from the current testing year; percentage proficient on the same content areas on Michigan Educational Assessment Program or MI-Access in elementary or middle school from the previous testing year; and four-year cohort graduation rates and four-year cohort dropout rates in high school.

Each of these performance indicators was standardized, creating school z-scores based on all schools reporting scores on that indicator. The performance measure was then created for each school by taking the mean of the z-scores over these performance indicators.

Specifically, the summary performance measure for each school (j) was calculated as follows:

$$Y_j = \frac{\sum_{i=1}^{N_j} z_{ij}}{N_j} \tag{B5}$$

where i is an indicator representing each outcome available for school j, N_j is the number of performance indicators available for school j, and z_{ij} is the z-score of school j on a specific performance indicator i.

Alternative performance measure. As discussed in the main report, this study used a study team–modified version of the initial (pre-2010/11) prediction method performance measures as an alternative to the Top-to-Bottom ranking. This alternative performance measure, applied to both methods, addresses some of the limitations of the pre-2010/11 measures. Specifically, the alternative differs from the pre-2010/11 measures in three aspects:

- *Use of weights.* The pre-2010/11 performance measures used for both methods were constructed from multiple school assessment results. In constructing these composite benchmark measures, assessment results were not weighted according to the number of students who took each test. For the alternative performance measure, assessment scores were weighted proportionally to the number of students who took each test.
- *Selection of assessments.* For the pre-2010/11 performance measure, the measure used by the prediction method was based on 12 assessments for the prediction method and up to 35 assessments for the measure used by the comparison method (depending on school type and student composition). As an alternative the study applied a common selection of core subjects (math and reading) for both methods.
- *Approach to aggregating multiple test results.* For the pre-2010/11 measure, the prediction method used a performance measure based on the individual z-values of assessment scores, while the comparison method used a performance measure based on school z-values of percent proficient on each assessment. As an alternative to applying different aggregation approaches to the two methods, the study used a common approach based on the prediction method.

In sum, the alternative performance measure is the (weighted) average of student *z*-scores based on core subject assessments and is a common measure for both methods.

Selection of school demographic characteristics

For beating-the-odds school identification, a set of school demographic characteristics was used as the covariates in the performance estimation in the prediction method and as the items on which to evaluate how similar schools are using the comparison method. As discussed in the main report, the study team adopted a different set of characteristics for each method as the benchmark specification, following Michigan's approach. For the prediction method, school demographic variables for students' free or reduced-price lunch, disability, and English language learner statuses are used. For the comparison method, in addition to low-income, disability, and English language learner indicators, other school characteristics (including locale, total enrollment, percent racial/ethnic minority enrollment, school configuration, special education center status, and state foundation allowance) are used.

As an alternative specification for both methods, the study selected the following common set of school characteristics: percent female students, total enrollment, percent English language learner students, percent economically disadvantaged students, percent students with disabilities, percent racial/ethnic minority students, grades served (elementary, middle, or high schools), magnet school indicator, and indicators for each locale. This alternative set of school characteristics was based on a series of stepwise multivariate regressions on the alternative composite performance measure, starting with all baseline school characteristic variables originally applied in both methods. The stepwise regressions were conducted separately using each of the four study years. Inputs that were significant at the 5 percent level were initially included in the performance estimation. The alternative set was selected by using criteria that the estimated coefficients are statistically significant for three out of four years as well as taking into account policy relevance. The select results from the stepwise regressions are presented in table B7 to illustrate the statistical significance of explanatory variables across years.

Estimation of identification models

As noted earlier, the prediction method involved the estimation of a prediction model (that is, the estimation of school performance) as a key feature. The results of the estimation of the prediction model under the alternative model (model E) for 2007/08–2010/11 are shown in table B7. The coefficient estimates characterize the prediction process for each year. The coefficient estimates included were significant for at least three of four years. Percent English language learner students and magnet school are exceptions—their inclusion is based on policy relevance.

The school identification results are illustrated in table B8. For the modified baseline model (model A') based on the pre-2010/11 Michigan-defined composite measure instead of the Top-to-Bottom measure, the table shows the average performance level in *z*-scores of schools that are identified as beating the odds, compared with that of schools not identified as beating the odds, and highlights how their performance levels differ from their corresponding demographically similar school clusters.

Table B7. Selection of school demographic characteristics for prediction method: regression results from stepwise regression models, by year

Characteristic	2010/11		2009/10		2008/09		2007/08	
	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Total enrollment	2.9E–04***	1.7E–05	2.7E–04***	1.8E–05	2.1E–04***	1.7E–05	2.0E–04***	1.7E–05
Percent female	1.072***	0.08	0.703***	0.08	0.488***	0.08	0.816***	0.08
Percent eligible for free or reduced-price lunch	–1.305***	0.02	–1.274***	0.02	–0.973***	0.02	–0.982***	0.02
Percent students with disabilities	0.283***	0.05	0.248***	0.05	0.207***	0.05	–0.011	0.05
Percent English language learner students	0.239***	0.07	0.210**	0.07	0.009	0.07	0.129	0.07
Percent racial/ethnic minority students	0.176***	0.05	0.201***	0.05	0.299***	0.05	0.244***	0.05
Serving grades K–5	0.071***	0.01	0.083***	0.01	0.057***	0.01	0.079***	0.01
Serving grades 6–8	–0.016	0.01	–0.022*	0.01	–0.061***	0.01	–0.039***	0.01
Serving grades 9–12	–0.241***	0.01	–0.270***	0.02	–0.307***	0.02	–0.319***	0.02
Magnet school	0.025	0.01	0.003	0.01	0.007	0.02	–0.009	0.02
Locale 1 (city, large)	–0.431***	0.03	–0.361***	0.03	–0.443***	0.03	–0.554***	0.03
Locale 2 (city, midsize)	–0.250***	0.03	–0.233***	0.03	–0.299***	0.03	–0.319***	0.03
Locale 3 (city, small)	–0.242***	0.03	–0.235***	0.03	–0.265***	0.03	–0.310***	0.03
Locale 4 (suburb, large)	–0.279***	0.02	–0.235***	0.03	–0.235***	0.03	–0.265***	0.03
Locale 5 (suburb, midsize)	–0.261***	0.03	–0.262***	0.04	–0.262***	0.04	–0.233***	0.04
Locale 6 (suburb, small)	–0.180***	0.04	–0.157***	0.04	–0.180***	0.04	–0.212***	0.04
Locale 7 (town, fringe)	–0.259***	0.04	–0.197***	0.04	–0.162***	0.03	–0.195***	0.03
Locale 8 (town, distant)	–0.200***	0.03	–0.165***	0.03	–0.140***	0.04	–0.206***	0.04
Locale 9 (town, remote)	–0.148***	0.03	–0.096**	0.03	–0.098**	0.04	–0.080*	0.04
Locale 10 (rural, fringe)	–0.203***	0.03	–0.166***	0.03	–0.143***	0.03	–0.169***	0.03
Locale 11 (rural, distant)	–0.160***	0.03	–0.122***	0.03	–0.099***	0.03	–0.125***	0.03
Constant	0.157**	0.05	0.278***	0.05	0.221***	0.05	0.055	0.05

* significant at the .05 level; ** significant at the.01 level; *** significant at the .001 level.

Note: The outcome variables are summary indexes based on *z*-scores.

Source: Authors' analysis based on Michigan Department of Education data.

The tables show that beating-the-odds schools are, on average, achieving a higher performance level than non-beating-the-odds schools and that their demographically similar school clusters in general performed at a lower level than non-beating-the-odds schools.

The main report provides the results from key comparisons of beating-the-odds school identification results using different technical specifications (that is, performance outcomes, school characteristics, and school configuration), statistical methods, and years examined. This section provides additional beating-the-odds school identification results and comparisons of the results across models that are not reported in the main report, including results for each model (models A–E), within-year analyses prior to 2010/11, comparisons with additional years, and comparisons using the performance measures Michigan Department of Education used prior to 2010/11. These additional comparisons were conducted to examine the robustness of the primary findings reported in the main report and present findings based on additional combinations of outcomes, characteristics, and configurations. The Top-to-Bottom performance outcome measure is limited to school

Table B8. Averages and standard deviations of performance measure for the comparison method, model A': beating-the-odds schools and non-beating-the-odds schools, by year

School year and performance level	Beating the odds schools		Non beating the odds schools	
	Number of schools	Mean	Number of schools	Mean
2007/08				
Average performance level	48	0.991	3,414	0.140
Average performance level of comparison group	48	0.112	3,414	0.180
Standard deviation of average performance of comparison group	48	0.274	3,414	0.298
2008/09				
Average performance level	51	0.849	3,386	0.103
Average performance level of comparison group	51	0.004	3,386	0.136
Standard deviation of average performance of comparison group	51	0.267	3,386	0.265
2009/10				
Average performance level	43	0.891	3,334	0.086
Average performance level of comparison group	43	0.065	3,334	0.116
Standard deviation of average performance of comparison group	43	0.253	3,334	0.254
2010/11				
Average performance level	58	0.922	3,221	0.080
Average performance level of comparison group	58	-0.026	3,221	0.110
Standard deviation of average performance of comparison group	58	0.266	3,221	0.248

Source: Authors' analysis based on Michigan Department of Education data.

year 2010/11 because that is the first year the newly constructed measure was used. No between-year comparisons could be made using the Top-to-Bottom measure.

Additional beating-the-odds school identification results: within-year, between-methods

Table 3 in the main report presents beating-the-odds school list agreement rates between the two statistical methods for 2010/11. The key finding is that the agreement rates between the two methods are not high, even when the methods are applied using comparable specifications.

Table B9 supplements table 3 by providing 2010/11 agreement rates between methods under baseline and alternative models, using Top-to-Bottom ranking for the baseline (model A) and for alternative models B, C, and D. The agreement rates are less than 50 percent for models B, C, D, consistent with the findings reported in table 3. The total number of schools identified by each method also varies by model, particularly when the outcome is changed from the baseline Top-to-Bottom ranking to the study team–developed alternative measure.

Table B10 supplements table 3 by providing 2007/08–2010/11 agreement rates between methods under the baseline and alternative models, with the baseline model using the Michigan-developed "initial" (pre-2010/11) measure. The agreement rates between the two methods as well as the number of schools vary by model and year. Of the agreements reported in table B10 across year and models, no pairwise comparison of the two methods had an agreement rate of greater than 50 percent. Consistent with the findings in the main report, table B10 shows that the identification results are unlikely to be very similar across the two methods.

Table B9. How do beating-the-odds school identification results vary by the statistical method used (research question 1)? Variation and agreement rate in school identification results across methods and models, 2010/11

Model specification	Baseline Model A		Model B		Model C		Model D		Model (C+D)	
Outcome	Top-to-Bottom		Alternative measure		Top-to-Bottom		Top-to-Bottom		Top-to-Bottom	
Characteristics	Michigan-selected		Michigan-selected		Alternative best fit		Michigan-selected		Alternative best fit	
Configuration	Pooled		Pooled		Pooled		Alternative (stratified by school)		Alternative (stratified by school)	
Method	Number of beating the odds schools	Agreement rate between methods (percent)	Number of beating the odds schools	Agreement rate between methods (percent)	Number of beating the odds schools	Agreement rate between methods (percent)	Number of beating the odds schools	Agreement rate between methods (percent)	Number of beating the odds schools	Agreement rate between methods (percent)
Prediction method	75	na	37	na	71	na	75	na	71	na
Comparison method	28	na	70	na	30	na	30	na	35	na
Both methods	20	39	22	41	18	34	19	38	17	32

na is not applicable.

Note: This table highlights the results from the between-method comparisons using 2010/11 measures. Model A (the baseline) reflects the current Michigan model. Comparison method results differ from Michigan's reported results because of differences in the treatment of missing data and weight applications in the Euclidean distance calculations. The weights used for each demographic characteristic in the comparison method in the table reflect those used by Michigan in 2010/11. In models B–D specifications on student outcomes, school characteristics, and sample configuration are altered one at a time while holding all other factors constant to baseline in order to gauge the influence of that factor on school identification. Model E incorporates alternative specifications for school characteristics and configuration but retains the Top-to-Bottom ranking as the outcome measure to maintain comparability of the results to current Michigan policy. The alternative measure has a similar construction to the performance measure for the prediction method in the initial Michigan model but instead weights scores by number of students tested, uses assessment scores from a common selection of core subjects, and bases the measure on individual-level *z*-values of assessment scores. In all models the same outcome measure is used for both methods given the selected outcome for that model. Best-fit school characteristics were selected based on a series of step-wise multivariate regressions on the alternative composite performance measure. For the alternative (stratified by school) configuration, schools were first separated into three groups: those serving elementary school grades (K–5), middle school grades (6–8), and high school grades (9–12). Beating-the-odds school identification was then conducted separately on each of the three groups.

Source: Authors' analysis based on Michigan Department of Education data.

Additional beating-the-odds school identification results: within-year, within-method

In the main report, table 4 presents beating-the-odds agreement rates when data and sample specifications are changed for 2010/11. The key finding is that the identification results do change when alternative specifications are applied and that changing the outcome measures from the Top-to-Bottom ranking to the study team–developed alternative performance index led to an agreement rate that is below 20 percent. As noted in the main report, the difference in these analytic samples between the two models compared might partly explain the variation in the identification results.

Table B11 extends the table 4 presentation of variation in identification results by model specification for 2010/11 by adding results for an alternative model that combines models C and D. Table B11 highlights, for each method, that the use of the alternative outcome measure led to a larger variation in the identification result from the baseline model than either the alternative school characteristics or the alternative school sample configuration, or the two combined.

Table B12 presents the variation in within-year, within-method identification results for 2007/08–2010/11 under the baseline and alternative models, with the baseline model using the Michigan-developed "initial" (pre-2010/11) measure. While table 5 in the main report

Table B10. How do the beating-the-odds school identification results vary by the method (research question 1)? Variation and agreement rate in school identification results across methods and models, 2007/08–2010/11

Model specifications	Initial Michigan model Model A		Model B		Model C		Model D		Model E' (pre 2010/11) and Model E (2010/11)	
Outcome	Michigan-defined composite performance measures		Alternative measure		Alternative measure		Alternative measure		Alternative measure	
Characteristics	Michigan-selected		Michigan-selected		Alternative best fit		Michigan-selected		Alternative best fit	
Configuration	Pooled		Pooled		Pooled		Alternative (stratified by school)		Alternative (stratified by school)	
	Number of beating the odds schools		Number of beating the odds schools		Number of beating the odds schools		Number of beating the odds schools		Number of beating the odds schools	
School year	Prediction	Comparison	Prediction	Comparison	Prediction	Comparison	Prediction	Comparison	Prediction	Comparison
2007/08	22	48	32	58	74	65	58	57	86	65
2008/09	27	51	35	60	72	68	63	62	85	78
2009/10	61	43	38	46	56	59	53	60	82	71
2010/11	42	58	37	70	65	73	51	74	71	80
School year	Number of schools identified in both methods	Agreement rate between methods (percent)	Number of schools identified in both methods	Agreement rate between methods (percent)	Number of schools identified in both methods	Agreement rate between methods (percent)	Number of schools identified in both methods	Agreement rate between methods (percent)	Number of schools identified in both methods	Agreement rate between methods (percent)
2007/08	8	23	17	38	25	36	21	37	34	45
2008/09	5	13	15	32	27	39	25	40	39	48
2009/10	7	14	13	31	20	35	19	34	29	38
2010/11	5	10	17	32	31	45	24	38	38	50

Note: The table highlights results from the within-year between-method comparisons using pre-2010/11 measures. The initial Michigan model (model A') uses the performance measures developed by Michigan (separately by method), Michigan-selected school characteristics, and school configuration. Comparison method results differ from Michigan's reported results because of differences in the treatment of missing data and weight applications in the Euclidean distance calculations. The weights used for each demographic characteristic in the comparison method in the table reflect those used by Michigan prior to 2010/11. The alternative measure has similar construction to the performance measure for the prediction method in the initial Michigan model but instead weights scores by number of students tested, uses assessment scores from a common selection of core subjects, and bases the measure on individual-level *z*-values of assessment scores. Best-fit school characteristics were selected based on a series of stepwise multivariate regressions on the alternative composite performance measure. Where the alternative measure is used (models B' and E'), the same measure is used for both methods. For the alternative (stratified by school) configuration, schools were first separated into three groups: those serving elementary school grades (K–5), middle school grades (6–8), and high school grades (9–12). Beating-the-odds school identification was then conducted separately on each of the three groups.

Source: Authors' analysis based on Michigan Department of Education data.

Table B11. How do the identification results vary when alternative performance measures and school sample configuration are used (research question 2)? Variation in beating-the-odds school identification results by model (by altering specification), based on the current Michigan model as baseline, by method, 2010/11

Model specification	Baseline model A	Model B			Model C			Model D			Model (C+D)		
Outcome	Top to Bottom	Alternative measure			Top to Bottom			Top to Bottom			Top to Bottom		
Characteristics	Michigan-selected	Michigan-selected			Alternative best fit			Michigan-selected			Alternative best fit		
Configuration	Pooled	Pooled			Pooled			Alternative (stratified by school)			Alternative (stratified by school)		
Method	Number of beating-the-odds schools	Number of beating-the-odds schools	Number common with baseline	Agreement rate (percent)	Number of beating-the-odds schools	Number common with baseline	Agreement rate (percent)	Number of beating-the-odds schools	Number common with baseline	Agreement rate (percent)	Number of beating-the-odds schools	Number common with baseline	Agreement rate (percent)
Prediction method	75	37	6	11	71	55	75	75	75	100	71	55	75
Comparison method	28	70	9	18	30	12	41	30	26	90	35	13	41

Note: The table highlights results from the within-year between-method comparisons using pre-2010/11 measures. The initial Michigan model (model A') uses the performance measures developed by Michigan (separately by method), Michigan-selected school characteristics, and configuration. Comparison method results differ from Michigan's reported results because of differences in the treatment of missing data and weight applications in the Euclidean distance calculations. The weights used for each demographic characteristic in the comparison method in the table reflect those used by Michigan prior to 2010/11. The alternative measure has similar construction to the performance measure for the prediction method in the initial Michigan model but instead weights scores by number of students tested, uses assessment scores from a common selection of core subjects, and bases the measure on individual-level *z*-values of assessment scores. Best-fit school characteristics were selected based on a series of stepwise multivariate regressions on the alternative composite performance measure. Where the alternative measure is used (models B' and E'), the same measure is used for both methods. For the alternative (stratified by school) configuration, schools were first separated into three groups: those serving elementary school grades (K–5), middle school grades (6–8), and high school grades (9–12). Beating-the-odds school identification was then conducted separately on each of the three groups.

Source: Authors' analysis based on Michigan Department of Education data.

Table B12. How do the identification results vary when alternative performance measures and school sample configuration are used (research question 2)? Variation in beating-the-odds school identification results by model (by altering specification), based on the initial Michigan model as prebaseline, by method and school year

Model specifications	Initial Michigan model Model A	Model B'			Model C'			Model D'			Model E'		
Outcome	Michigan-defined composite performance measures	Alternative measure			Alternative measure			Alternative measure			Alternative measure		
Characteristics	Michigan-selected	Michigan-selected			Alternative best fit			Michigan-selected			Alternative best fit		
Configuration	Pooled	Pooled			Pooled			Alternative (stratified by school)			Alternative (stratified by school)		
Method and school year	**Number of beating-the-odds schools**	**Number of beating-the-odds schools**	**Number common with pre-baseline**	**Agreement rate (percent)**	**Number of beating-the-odds schools**	**Number common with pre-baseline**	**Agreement rate (percent)**	**Number of beating-the-odds schools**	**Number common with pre-baseline**	**Agreement rate (percent)**	**Number of beating-the-odds schools**	**Number common with pre-baseline**	**Agreement rate (percent)**
Prediction method													
2007/08	22	32	17	63	74	17	35	58	17	43	86	17	32
2008/09	27	35	19	61	72	22	44	63	23	51	85	24	43
2009/10	61	38	25	51	56	23	39	53	28	49	82	28	39
2010/11	42	37	21	53	65	21	39	51	22	47	71	24	43
Comparison method													
2007/08	48	58	24	45	65	19	34	57	25	48	65	17	30
2008/09	51	60	20	36	68	16	27	62	16	28	78	19	30
2009/10	43	46	13	29	59	13	26	60	16	31	71	16	28
2010/11	58	70	18	28	73	14	21	74	19	29	80	15	22

Note: The table highlights results from the within-year, within-method comparisons using pre-2010/11 measures. The initial Michigan model (model A') uses the performance measures developed by Michigan (separately by method), Michigan-selected school characteristics, and configuration. Comparison method results differ from Michigan's reported results because of differences in the treatment of missing data and weight applications in the Euclidean distance calculations. The weights used for each demographic characteristic in the comparison method in the table reflect those used by Michigan prior to 2010/11. The alternative measure has similar construction to the performance measure for the prediction method in the initial Michigan model but instead weights scores by number of students tested, uses assessment scores from a common selection of core subjects, and bases the measure on individual-level *z*-values of assessment scores. Best-fit school characteristics were selected based on a series of stepwise multivariate regressions on the alternative composite performance measure. Where the alternative measure is used (models B' and E'), the same measure was used for both methods. For the alternative (stratified by school) configuration, schools were first separated into three groups: those serving elementary school grades (K–5), middle school grades (6–8), and high school grades (9–12). Beating-the-odds school identification was then conducted separately on each of the three groups.

Source: Authors' analysis based on Michigan Department of Education data.

displays how the agreement rate changes when specifications are altered one at a time, table B12 provides agreement rates under various combinations of alternative specifications. Different combinations of alternative specifications are presented in table B12 to allow observation of additional variation patterns such as:

- As described earlier, the study team–developed alternative measure is based on the Michigan-defined (pre-2010/11) outcome measures, while the Top-to-Bottom ranking measure is conceptually and mechanically different from the Michigan-defined (pre-2010/11) outcome measures. As might be expected, for each method, the difference between the identification results of the models using the alternative measure (B') and the identification results of the Michigan-defined composite performance measure (A') is smaller (that is, agreement rates are higher) than the difference between the results when using the alternative measure (B) and the alternative measure (A) reported in table 4. However, the agreement rates between the models using the alternative measure and Michigan-defined (pre-2010/11) measures are still 28–63 percent, indicating the sensitivity of the results to the choice of performance measures.
- As with table B11, table B12 shows that adding alternative specifications would lower the agreement rates. Table B12 demonstrates this pattern for alternative school characteristics or school configuration, in addition to the alternative measure (B' vs. C'; B' vs. D'; B' vs. E').

Additional beating-the-odds school identification results: between-years, within-methods

Table 5 in the main report presents beating-the-odds school agreement rates between two adjacent school years: 2007/08 and 2010/11. The key finding is that average year-to-year agreement rates over the four-year period did not exceed 50 percent for either method.

Table B13 extends table 5 by presenting between-years, within-methods agreement rates under various other sets of model specifications. The last column (E) of table B13 repeats the information reported in table 5, while models A' to D' show the results under alternative data and sample specifications. Consistent with table 5, table B13 shows that the agreement rates between any adjacent years is less than 60 percent for all reported models, with average rates of no greater than 50 percent. These observations are consistent with the findings reported with table 5 and indicate that each method is sensitive to the change in school performance underlying the input data.

Table B13. How do the school identification results vary from year to year (research question 3)? Variation in beating-the-odds school identification results across years, initial Michigan model and alternative models: number and ratio of matched beating-the-odds schools between years, within method

Model specifications		Initial Michigan model Model A'		Model B'		Model C'		Model D'		Model E (2010/11)	
Outcome		Michigan-defined composite performance measures		Alternative measure		Alternative measure		Alternative measure		Alternative measure	
Characteristics		Michigan-selected		Michigan-selected		Alternative best fit		Michigan-selected		Alternative best fit	
Configuration		Pooled		Pooled		Pooled		Alternative (stratified by school)		Alternative (stratified by school)	
Period 1	**Period 2**	**Number of schools identified in both periods**	**Agreement rate across periods (percent)**	**Number of schools identified in both periods**	**Agreement rate across periods (percent)**	**Number of schools identified in both periods**	**Agreement rate across periods (percent)**	**Number of schools identified in both periods**	**Agreement rate across periods (percent)**	**Number of schools identified in both periods**	**Agreement rate across periods (percent)**
Prediction method											
2007/08	2008/09	6	25	9	27	29	40	23	38	36	42
2008/09	2009/10	9	21	16	44	28	44	27	47	43	52
2009/10	2010/11	18	35	15	40	25	41	24	46	38	50
Average agreement rate		na	27	na	37	na	42	na	44	na	48
Comparison method											
2007/08	2008/09	23	47	31	53	31	47	35	59	34	48
2008/09	2009/10	14	30	23	43	32	50	30	49	34	46
2009/10	2010/11	16	32	30	52	33	50	29	43	40	53
Average agreement rate		na	36	na	49	na	49	na	50	na	49

na is not applicable.

Note: The table highlights results from the between-years, within-method comparisons using pre- 2010/11 measures. The initial Michigan model (model A') uses the performance measures developed by Michigan (separately by method), Michigan-selected school characteristics, and configuration. Comparison method results differ from Michigan's reported results because of differences in the treatment of missing data and weight applications in the Euclidean distance calculations. The weights used for each demographic characteristic in the comparison method in the table reflect those used by Michigan prior to 2010/11. The alternative measure has similar construction to the performance measure for the prediction method in the initial Michigan model but instead weights scores by number of students tested, uses assessment scores from a common selection of core subjects, and bases the measure on individual-level *z*-values of assessment scores. Best-fit school characteristics were selected based on a series of stepwise multivariate regressions on the alternative composite performance measure. Where the alternative measure was used (models B' and E'), the same measure was used for both methods. For the alternative (stratified by school) configuration, schools were first separated into three groups: those serving elementary school grades (K–5), middle school grades (6–8), and high school grades (9–12). Beating-the-odds school identification was then conducted separately on each of the three groups.

Source: Authors' analysis based on Michigan Department of Education data.

Notes

The authors thank the colleagues and partners who made the study possible from the early design phases to the final analysis. They thank the members of Michigan's Beating the Odds Research Alliance, including the Michigan Department of Education staff, research staff from two intermediate school districts, and leadership from Learning Forward for their input and guidance for the study. They especially thank Vanessa Keesler, Deputy Superintendent of Education Services at the Michigan Department of Education, for her continued support and guidance. Keesler's team, particularly Alexander Schwartz, provided the data as well as invaluable technical assistance. Fannie Tseng, formerly of Berkeley Policy Associates, made important early contributions to the design of the study. They also thank Rebecca Herman and Ayrin Molefe.

1. Schools identified as performing better than expected have been referred to as "beating-the-odds schools," "high-flying schools," or some derivation of "high-performing/high-poverty schools." This report uses the term "beating the odds" for consistency with several recent, ongoing state and local initiatives.
2. For school years 2009/10 and 2010/11, Michigan identified 184 schools as beating the odds by one or both of the two methods and in at least one of the two academic years. In 2009/10, 109 schools were identified as beating the odds by at least one methodology, and 26 schools were identified by both. In 2010/11, 121 schools were identified by at least one methodology, and 20 schools were identified by both. Only 46 schools were identified by either methodology for consecutive years, and only 4 of the 184 schools were identified by both methodologies in both years.
3. The number of overlaps may increase if the criteria used to identify beating-the-odds schools for each method are relaxed and the number of schools identified increases.
4. For the study, 50 separate models were estimated (25 per method), including models that applied pre-2010/11 performance measures used by the Michigan Department of Education.
5. Michigan Educational Assessment Program and MI-Access assessments are administered in the fall for the full previous school year of instruction. For example, for 2007/08, students are tested in fall 2008.
6. Prior to 2010/11, total enrollment was used instead.
7. For the results to be comparable from year to year, a separate analysis was conducted using the weights used prior to 2010/11, when the Top-to-Bottom ranking was not available. The weights, provided by the Michigan Department of Education, were: 6 for percentage of students eligible for free or reduced-price lunch, 2 for percentage students with disabilities, 10 for being a special education center, 2 for state foundation allowance, 2 for total enrollment, 12 for each locale, 13 for each school configuration indicator, and 1 otherwise.
8. Since 2010/11, Michigan has selected the 29 most similar schools as a comparison school group. Prior to that, Michigan selected the 30 most similar schools. Following Michigan's approach, the study team selected the 30 closest schools for pre-2010/11 and the 29 closest schools for 2010/11. When year-to-year comparisons are made, however, the 30 closest schools were selected for 2010/11 to allow comparisons across all years.
9. The agreement rate R_{ij} relates similarly to a commonly used Jaccard's index: $S_{ij} = n_{ij}/(N_i + N_j - n_{ij})$ by a factor of $(2 - R_{ij})$.
10. For the Michigan Educational Assessment Program and MI-Access, only math and reading scores are used. The Michigan Educational Assessment Program writing section was different prior to 2009/10. Scores were included where available.

References

Apthorp, H., Englert, K., Barley, Z., Lauer, P., Gamache, L., Buhler, B. V., et al. (2005). *Final report: High-needs schools—What does it take to beat the odds?* Aurora, CO: Mid-Continent Research for Education and Learning. Retrieved May 8, 2012, from http://www.mcrel.org/~/media/Files/McREL/Homepage/Products/01_99/prod57_BeatTheOdds.ashx.

Ascher, C., & Maguire, C. (2007). *Beating the odds: How thirteen NYC schools bring low-performing ninth graders to timely graduation and college enrollment.* Providence, RI: Annenberg Institute for School Reform at Brown University. Retrieved July 17, 2012, from http://annenberginstitute.org/pdf/BTO_report.pdf.

Carter, S. C. (2000). *No excuses: Lessons from 21 high-performing, high-poverty schools.* Washington, DC: Heritage Foundation. http://eric.ed.gov/?id=ED440170

Connell, N. (1999). *Beating the odds: High-achieving elementary schools in high-poverty neighborhoods.* New York, NY: Education Priorities Panel. http://eric.ed.gov/?id=ED441913

Cudeiro, A., Palumbo, J., & Nelsen, J. (2005). *Six schools that make a difference.* Sacramento, CA: Association of California School Administrators.

The Education Trust. (2005). *Gaining traction, gaining ground: How some high schools accelerate learning for struggling students.* Washington, DC: The Education Trust. Retrieved January 30, 2012, from http://www.edtrust.org/sites/edtrust.org/files/publications/files/GainingTractionGainingGround.pdf.

Grusenmeyer, L., Fifield, S., Murphy, A., Nian, Q., & Qian, X. (2010). *Delaware middle schools: Beating the odds.* Newark, DE: University of Delaware, Delaware Education Research and Development Center. Retrieved May 8, 2012, from http://udspace.udel.edu/bitstream/handle/19716/5532/Delaware%20Middle%20Schools%20Beating%20the%20Odds_T2010.4.pdf?sequence=1.

Harris, D. (2007). High-flying schools, student disadvantage, and the logic of NCLB. *American Journal of Education, 113,* 367–394. http://eric.ed.gov/?id=EJ757581

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). *Turning around chronically low-performing schools* (IES Practice Guide, NCEE 2008–4020). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved January 25, 2012, from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/Turnaround_pg_04181.pdf.

Kearney, W., Herrington, D., & Aguilar, D. (2012). Beating the odds: Exploring the 90/90/90 phenomenon. *Equity & Excellence in Education, 45*(2), 239–249.

Langer, J. (2000). *Beating the odds: teaching middle and high school students to read and write well* (2nd ed.). New York: National Research Center on English Learning and

Achievement, University at Albany, State University of New York. Retrieved November 16, 2012, from http://www.albany.edu/cela/reports/langer/langerbeating12014.pdf.

Pérez, M., Anand, P., Speroni, C., Parrish, T., Esra, P., Socías, M., et al. (2007). *Successful California schools in the context of educational adequacy*. Palo Alto, CA: American Institutes for Research. http://eric.ed.gov/?id=ED499199

Reeves, D. (2004). *Accountability in action: A blueprint for learning organizations*. Denver, CO: Advanced Learning Press.

Sebring, P. B., Allensworth, E., Bryk, A., Easton, J. Q., & Luppescu, S. (2006). *The essential supports for school improvement*. Chicago, IL: Consortium on Chicago School Research, University of Chicago. Retrieved January 25, 2012, from http://ccsr.uchicago.edu/sites/default/files/publications/EssentialSupports.pdf.

Socias, M., Dunn, L., Parrish, T., Muraki, M., & Woods, L. (2007). *California high schools that beat the odds in high school graduation*. Santa Barbara, CA: California Dropout Research Project. Retrieved July 17, 2012, from http://www.hewlett.org/uploads/files/California_High_Schools_That_Beat_the_Odds_in_High_School_Graduation.pdf.

Taylor, B., Pearson, P., Clark, K., & Walpole, S. (1999). *Beating the odds in teaching all children to read*. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. Retrieved April 23, 2012, from http://www.ciera.org/library/reports/inquiry-2/2-006/2-006.pdf.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2009). *Common Core of Data: Local Education Agency Universe Survey, 2007–08*. Retrieved April 1, 2012, from http://nces.ed.gov/ccd/ccddata.asp.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2010). *Common Core of Data: Local Education Agency Universe Survey, 2008–09*. Retrieved April 1, 2012, from http://nces.ed.gov/ccd/ccddata.asp.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2011). *Common Core of Data: Local Education Agency Universe Survey, 2009–10*. Retrieved April 1, 2012, from http://nces.ed.gov/ccd/ccddata.asp.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2012). *Common Core of Data: Local Education Agency Universe Survey, 2010–11*. Retrieved April 1, 2012, from http://nces.ed.gov/ccd/ccddata.asp.

Waits, M. J., Campbell, H. E., Gau, R., Jacobs, E., Rex, T., & Hess, R. K. (2006). *Why some schools with Latino children beat the odds…and others don't*. Tempe, AZ: Morrison Institute for Public Policy. Retrieved May 8, 2012, from http://www.arizonafuture.org/research/pdf/FAZ502_LatinEd_final.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