

**National Center for
Educational
Accountability**

national sponsor of Just  for the Kids

Identifying and Studying High-Performing Schools

NCEA Issue Brief # 1 – January 2006

Chrys Dougherty, Ph.D., National Center for Educational Accountability

Identifying and Studying High-Performing Schools

Chrys Dougherty, Ph.D., National Center for Educational Accountability¹

Part 1: Identification of High-Performing Schools

Researchers and education leaders may seek to identify high-performing schools for at least three purposes. The first is to study how the practices in high-performing schools differ from those in average- or low-performing schools, in the hope that some of these practices will prove useful in improving other schools. The second is to use the higher performing schools' accomplishments to convince educators in other schools that they can do significantly better with "their kinds of kids." The third purpose is to publicly honor and reward high-performing schools to motivate others to emulate the desirable practices in those schools.

Since a major object of the study of high-performing schools is to discover practices that are worth emulating, the methodology used to identify those schools should meet several requirements:

- The analysis must use enough data to distinguish sustained high performance or improvement from that resulting from random year-to-year changes in the data.
- The analysis should distinguish three types of cases: 1) sustained high performance, 2) improvement to high performance by initially average-performing schools; and 3) improvement to average or high performance by initially low-performing schools. All three types of cases are worth studying to differentiate the practices that are present in each.
- The analysis should also distinguish cases in which higher performing or improving schools are nested in higher performing or improving districts from cases where they are not, in order to facilitate investigation of the role of the district in school improvement.
- The areas in which the school is high performing must be clearly identified. For example, a school might be high performing in reading and mathematics but not science; in getting poorly prepared students up to minimum standards but not in getting better prepared students up to advanced standards; in Advanced Placement exam success by affluent students but not by low-income students; and so on.
- The indicators chosen to measure school performance should reflect an explicit set of goals. For example, if the goal is to promote advanced academic performance, indicators that look only at minimum skills achievement would probably not be satisfactory.²

¹ The author would like to thank Veronica Brinson, Doug Carnine, Chrys Dougherty III, Jean Rutherford, Jerry Silbert, Susan Whisenant, and Heather Zavadsky for their kind and helpful comments on this paper. Any mistakes in the article are my own.

² By similar reasoning, a researcher or policymaker interested in students' socioemotional development might want to use a survey on student behavior rather than achievement test results to define success.

Researchers will of course be aware that correlates of high performance are not necessarily causal, and that research must still be conducted on whether practices found more often in high-performing schools a) can be introduced systematically into average- or low-performing schools, and b) improve those schools' performance when introduced.³

Types of School Performance Measures

For a given set of school effectiveness indicators, the measures and criteria used to identify consistently high-performing schools may be divided into three types:

- 1. Relative performance measures:** These measures identify schools that are more successful than other schools serving similar students: Do the school's students perform better than would be predicted by their demographics and prior achievement, at a level that is unlikely by random chance?⁴ On the condition that the analysis is able to control for between-school differences in the most important non-school influences, measures of relative school performance can be interpreted as rough measures of school effectiveness.
- 2. Consistency measures:** Success should be examined over multiple years to make sure that the school is not being identified based on random year-to-year changes in student cohorts. Educators and policymakers are also interested in success that is reasonably consistent across student subpopulations: a high-performing school should not have any consistently low-performing student groups.
- 3. Absolute performance measures:** Relative performance criteria may fail to emphasize that performance by disadvantaged students needs to improve, even in the relatively more successful schools serving those students. For that reason and in order not to convey a message that less is expected of low-income students, state accountability systems and Adequate Yearly Progress (AYP) ratings generally rely on absolute measures. Since these measures give an automatic advantage to schools serving advantaged student populations, they should not be thought of as school effectiveness measures, but rather as checks that students in the schools identified as relatively higher performing also meet minimum standards of absolute performance.

Each of these types of performance measures has its counterpart in the world of improvement: improvement in relative performance (the school improved faster than similar schools); improvement in absolute performance (the school improved, but not necessarily faster than similar schools); and improvement in consistency of performance. In the following section we discuss the three types of performance measures in more detail.

³ Random selection of schools in which to apply the intervention in question may be the closest researchers can come to a "clinical trials" approach to testing out the effectiveness of such interventions.

⁴ Relative performance measures are a form of "norm-referencing for schools" in that they compare schools' performance with that of other schools, without regard to the absolute level that that performance represents.

Relative performance measures

Relative performance measures are typically based on a regression model that attempts to identify how the school performs given the population that it serves, and how a school performs with a given subpopulation given the characteristics of that subpopulation. The following ideas should be kept in mind when constructing these models.

Use of longitudinal student data makes it possible to control for prior student achievement and to identify how long students have been enrolled in a particular school. This is particularly important for identifying the “value-added” of teachers in a particular grade, and for school effectiveness analysis at the middle and high school levels. For elementary grades where no prior year test data are available, the analysis can give more weight to the performance of students that have been enrolled in the school for longer and whose achievement thus better reflects the impact of the school’s instructional program.⁵

Separate estimates of effectiveness by grade, subject, and student group make it possible to examine the consistency in performance across grades and subjects within the same school. Multiple years of data should be used to begin to separate school effects from random year-to-year fluctuations. The analysis should also allow for the greater variability and year-to-year volatility of results for small student groups.

Use of multiple indicators can make it possible to examine the consistency of the school’s performance across those indicators. At the high school level, these indicators can include college-ready graduation rates⁶ and population passing rates on Advanced Placement and International Baccalaureate exams.⁷

Separate estimates of effectiveness for students disaggregated by prior achievement levels makes it possible to identify schools that do a good job of producing growth with well-prepared and ill-prepared students. Whether a student is academically well prepared on entry into the school or classroom may be the single most educationally relevant piece of background information about that student.

Separate analysis at the classroom level can make it possible to examine a) the distribution of teacher effects within a school; b) how much of the variance in student performance is between classrooms within the same school vs. between schools; and c) the extent to which high-performing schools tend to outperform average- and low-performing schools because they have

⁵ Because the use of longitudinal student data is so important for school and teacher effectiveness analysis, NCEA has launched a nationwide Data Quality Campaign in conjunction with nine other organizations to promote statewide longitudinal data systems and availability of longitudinal student data to researchers under appropriate privacy protections. Achieve and NCEA have also released a publication, *Creating a Longitudinal Data System: Using Data to Improve Student Achievement*, outlining Ten Essential Elements of a statewide longitudinal student data system. For a downloadable copy of the publication and a list of the organizations involved in the campaign, see www.dataqualitycampaign.org.

⁶ College-ready graduation rates are defined as the percent of a longitudinal student cohort who both graduate from high school and meet college readiness standards on state or national exams, such as the SAT or ACT.

⁷ Population exam passing rates use the percent of the *entire student cohort* who take and pass AP or IB exams, not just the percent of AP or IB *exam-takers* who pass. The latter rate may be inflated by discouraging all but a few top students from taking the exams.

more teachers at the top of the distribution or fewer teachers near the bottom. However, data linking individual students to their teachers is often difficult to obtain.⁸

Sensitivity analysis should be conducted to determine the extent to which “small” changes in the analytical model can lead to large changes in calculated effectiveness measures and in the derived list of schools identified as relatively more effective.

Criteria for preferring one statistical model to another should be developed. For example, models might be compared based on their predictive power: Model A is considered to be better than Model B if, when the two models are applied over three years of data leading up to time t , Model A is better at predicting whether a school will be high performing in time $t+1$. (This is especially true if Model A predicts $t+1$ performance better even when that performance is measured according to Model B.) Different models may also be more suited to answering different research questions.

Differentiation of “regular” schools from magnet schools with selective admissions criteria makes it possible to identify schools that “beat the odds” without having the advantage of serving selective student populations.

Consistency Measures

Consistency analysis looks at a) how schools that are high performing on average across grades or student groups perform with each grade or group separately; and b) the extent to which high-performing schools on Indicator A are also high performing on Indicator B. Consistency may be approached in several ways. One is to investigate the consistency of the measures themselves, by calculating statistics on the consistency of school performance across those measures. These statistics may include the variance in school performance ranks across measures; the correlation in ranks between pairs of measures; and the distribution of schools’ differences in ranks across measures.

A second approach is to give schools a “consistency score” in addition to a “performance score.” Schools with lower cross-indicator variance in performance than would be predicted by the size of the groups being analyzed are more consistent in their performance. Alternatively, relative performance below a certain minimum threshold (say, in the bottom half of schools) may be counted as a “miss,” and a school’s percentage of misses may be compared with what would be expected from a school with the same student group sizes and school average performance.

Absolute Performance Measures

Absolute performance measures are generally used to review whether schools that perform well on relative performance measures also reach minimum acceptable levels of performance (“performance floors”) on absolute measures. Examples of performance floors include AYP criteria and the performance measures used to determine accountability ratings in many states.

⁸ The ability to match teachers to students is one of the Ten Essential Elements of a student data system identified by the Data Quality Campaign (see footnote 5 above). In a summer 2005 survey by NCEA, thirteen states reported that their data systems would have the ability as of the 2005-06 school year to match individual students to teachers.

Part 2: Research on Practices in High-Performing Schools

Researchers may want to investigate how the practices found in high-performing schools vary systematically from those of schools that are average or low performing on the same indicators. Schools that are high performing on Indicator A but not B may also be contrasted with other schools that are high performing on both indicators. The purpose of these studies is to a) identify practices that are worthy of further study, and b) provide educators with specific information on “how high-performing schools do it” that can be helpful in their own improvement efforts.⁹

Researchers designing such studies of “best practices” or “high-performing school practices” will want to keep several considerations in mind:¹⁰

- 1. The research should be based on a conceptual framework.** For example, the National Center for Educational Accountability has developed the NCEA Best Practice Framework as a taxonomy for organizing the study of school practices.¹¹ A good conceptual framework facilitates not only research but also explanation and dissemination of the practices in question.
- 2. The research methodology should be carefully tailored to the research questions to be answered.** For example, research on the most effective ways of teaching elementary school students how to add fractions might use a different methodology than research investigating how teachers and administrators in the school decide on how best to teach fractions. Research on effective methods of teaching fractions might include cognitive science research on the understandings students must have to master the skills and teachers to teach them; experimental studies in which different teaching approaches are tried in experimental and control classrooms; and observational studies to differentiate the methods actually used in those classrooms. Studies of how educators make decisions on how to teach fractions might be based on interviews and surveys, with care taken to examine the consistency of answers given by different educators.
- 3. The division of labor between district, school administration, and classroom needs to be studied.** Evidence from the study of high-performing schools indicates that some functions, such as development of clear, specific, and vertically aligned academic standards and benchmark assessments, can best be handled at the district level, while other functions, such as selection of suitable professional development for a school’s teachers, may best be handled at the school level.¹²
- 4. Practices are best researched when they are defined as specific behaviors, e.g., the frequency of teacher team meetings and the content of discussions in those meetings, rather**

⁹ The assumptions here are that a) this information will be more valuable than most of the advice to which educational practitioners are currently exposed, and b) research based on experimental designs is not currently available for most educational questions, and indeed many of these questions do not lend themselves to such designs.

¹⁰ The label “best practices” is often applied not because the practices have been proven to be the very “best,” but because they are practices that are supported as beneficial by the best currently available evidence.

¹¹ See www.just4kids.org/bestpractice/framework.cfm?sub=framework for a description of the NCEA Best Practice Framework.

¹² The district’s role has been rediscovered by charter schools, whose consistency of quality is more easily maintained if they are part of a “chain” with its own charter management organization.

than just, “Do teachers collaborate with their colleagues?” The self-audits developed by NCEA provide examples of those behaviors.¹³

5. **Practices should not be looked at in isolation.** Interaction effects are likely to be common; for example, the effectiveness of a newly introduced professional development strategy is likely to depend on its relationship to the curriculum being taught.
6. **Practices in schools improving from low or average performance should be treated as a separate topic from practices in schools with sustained high performance.** Practices in schools with sustained high performance may differ from those in improving schools. Practices in improving schools may vary based on where a school is on its improvement trajectory and where the starting point was.
7. **Analysis must be conducted to learn how sensitive the findings are to changes in the research methodology.** For example, the number and identity of persons being interviewed, the coding of interview results by different researchers, the types of documentary evidence collected, and the wording of survey items all introduce sources of variability that must be investigated.
8. **The number of schools and interviewees needed to produce reliable results must also be investigated.** Many studies rely on research in a handful of schools. This is related to the sensitivity analysis discussed above: How sensitive are the results to interviewing different people or visiting a different set of (equally eligible) schools?

Conclusion

The identification of high-performing or “beating the odds” schools can have a powerful influence in convincing educators that greater success is possible with the students in their care. This may help to change the mindset of educators from one of “meeting minimum accountability standards in order to stay out of trouble” to that of “pushing out the boundaries of what is possible with our students.” The analysis of beating the odds schools must carefully identify the areas of high performance, the consistency of that performance, and the extent to which those schools outperform their average-performing counterparts. Magnet schools that select highly motivated students are likely to have higher student performance, but the more interesting schools to study are those that “beat the odds” without the advantage of serving selective student populations.

The study of practices in “beating the odds” schools must be carefully designed so that the research questions are carefully defined and the findings are not overly sensitive to changes in research methodology and personnel. In addition, practices and clusters of practices found to be associated with high performance need to be investigated further, by studying the effects of interventions designed to bolster those practices in average- and low-performing schools.¹⁴

¹³ See www.just4kids.org/bestpractice/self_audit_framework.cfm?sub=tools for the content of the NCEA self-audits.

¹⁴ Improvement strategies found to be effective in helping low-performing schools to become average performing may be different from those that help average-performing schools become high performing.