



### About Education Service Providers

For the purposes of this report, *ESP* is defined as a nonprofit or for-profit organization that contracts with new or existing public, charter, or private schools and/or school districts to provide comprehensive services to schools, including, but not limited to, educational programming and administrative services. Educational programming includes curriculum design, professional development, and student assessment tools. Administrative services include, but are not limited to, operation management services (student enrollment, school marketing), financial management services (payroll assistance, budget oversight), facilities management services (maintenance and use of facilities), and human resources management (hiring and training staff, staff benefits). The comprehensive services provided by ESPs are comparable in many ways to services provided by more traditional whole school improvement providers, such as the comprehensive school reform models that the CSRQ Center has reviewed in previous reports (<http://www.csrq.org>).

- Category 2: Evidence of positive effects on additional outcomes
- Category 3: Evidence of positive effects on parent, family, and community involvement
- Category 4: Evidence of link between research and the model's design
- Category 5: Evidence of services and support to schools to enable successful implementation:
  - a. Evidence of readiness for successful implementation
  - b. Evidence of professional development/technical assistance for successful implementation

This consumer guide provides education stakeholders with a decision-making tool that can help them to sort through the range of ESP models that are available to support whole school or district improvement. The guide also allows education consumers to compare individual ESPs with traditional comprehensive school reform (CSR) models that the CSRQ Center has reviewed in previous reports. As with all reports from the CSRQ Center, these reviews are intended to clarify options, not to point to or endorse best buys from among the seven ESP models that are profiled.

### The CSRQ Center's Findings

Our rating process for Categories 1, 2, and 3 is complex and combines two elements to provide a single rating:

- The strength of the evidence based on the causal validity of the research design (e.g., how reliable and credible is it)
- The strength of the reported impact or effect (e.g., does the model raise student achievement a little or a lot)

The CSRQ Center identified few rigorous studies that were relevant for rating each ESP's overall evidence of positive effects on student achievement. In Category 1, after screening approximately 900 studies and documents for quality, only 9 studies met the CSRQ Center's standards. (Table 1 summarizes the quantitative study findings that were

### About the CSRQ Center's Rating System

The production of this report was guided by the CSRQ Center's Quality Review Tool (QRT). The QRT provides criteria and procedures for scientifically rigorous, independent, fair, and credible model reviews. (For more details, see "Methodology.") Using the QRT, the CSRQ Center arrived at its ratings by applying separate rubrics to each of the five categories of effectiveness and quality. The ratings are expressed by a set of symbols:

- Very Strong ● . This is the highest rating provided by the CSRQ Center.
- Moderately Strong ◐ . This is the second highest rating.
- Moderate ◑ . Models that received this rating may still have notable evidence of positive outcomes but they may not be as strong as those that received a very strong or moderately strong rating.
- Limited ◒ . This rating indicates that while the CSRQ Center found some evidence of effectiveness, more rigorous research needs to be conducted to produce evidence to fully support the model's effectiveness or quality in the category or subcategory reviewed.
- Zero ◓ . This rating means that while the CSRQ Center found evidence that allowed it to provide a rating for a category or subcategory, none of the studies were of sufficient quality to be counted as reliable evidence.
- Negative ⊖ . This rating indicates that the CSRQ Center found strong evidence of detrimental effects in a given category or subcategory. In practice, we did not find any evidence of this kind for any model.
- No Rating (NR) . This rating indicates that a model had no studies (i.e., evidence) available for review in a category or subcategory.

used to rate evidence of overall positive effects on student achievement.) These 9 studies focused on one ESP model (Edison).

For Category 1 (Evidence of Positive Effects on Student Achievement), we rated

- One model as moderate (Edison),
- Four models as zero (Leona, Mosaica, NHA, and White Hat), and
- Two models as no rating (Imagine and SABIS).

In reviewing our findings for Category 1, readers should keep in mind that many of the models in the report serve high-poverty students in low-performing schools. Thus, the evidence of effectiveness that the studies present is for success in educating students in highly challenging conditions.

The research base on which to rate models in Categories 2 (Evidence of Positive Effects on Additional Outcomes) and 3 (Evidence of Positive Effects on Parent, Family, and Community Involvement) is sparse. There were no studies that met the CSRQ Center's standards in these categories.

Category 4 rated evidence of link between research and the model's design. The rating system for Category 5 (Evidence of Services and Support to Schools to Enable Successful Implementation) depended on two subcategories: (a) evidence of readiness for successful implementation and (b) evidence that the model provider offers professional development and technical assistance to enable successful implementation. The same rating scale and symbols were used to rate Categories 4 and 5 as were used to rate Categories 1–3; however, the meanings of the ratings are category specific. The CSRQ Center contacted each ESP to conduct a conversation with the provider to verify information to complete ratings in Categories 4 and 5. However, four models (Imagine, Mosaica, SABIS, and White Hat) did not engage in conversations with the CSRQ Center and three of those models (Imagine, Mosaica, and White Hat) did not provide any information. Thus, given the limited amount of publicly available information in these categories, we assigned no rating to these ESPs for Categories 4 and 5.

For Category 4, we rated

- One model as very strong (Edison),
- One model as limited (Leona), and
- One model as zero (NHA).

For Category 5a (evidence of readiness for successful implementation), one model (Edison) was rated very strong and two models (Leona and NHA) were rated moderate. For Category 5b (evidence that the ESP offers professional development and technical assistance to enable successful implementation), one model (Edison) was rated very strong and two models (Leona and NHA) were rated moderately strong.

Given the importance of implementation to the success of any whole school reform, consumers who select models that have low rankings in evidence of effects on student outcomes may still experience success if they implement the models faithfully. Table 2 summarizes basic model information and model ratings for Categories 1–5.

### Background on Whole School Improvement

For the past decade, increased attention has focused on finding effective strategies to improve schools and student achievement. Whole school improvement, sometimes referred to as CSR, is one approach. It has been tried in thousands of schools nationwide, mostly for high-poverty students in low-performing schools.

This trend is driven by the recognition that school improvement efforts are complex and require a coordinated, systematic approach that addresses every aspect of a school, including curriculum; instruction; governance; scheduling; professional development; assessment; and parent, family, and community involvement. Rather than use individual, piecemeal programs or approaches, effective whole school improvement is meant to integrate research-based practices into a unified effort to raise student achievement and achieve other important outcomes, such as reducing dropout rates or improving behavior. To support implementation, whole school improvement models typically provide schools with materials, professional development, and technical assistance.

Given the increased emphasis on accountability for results that was established by NCLB, education decision makers are continually searching for new and more effective school improvement options. In this context, alternatives to

traditional approaches—such as charter schools and ESPs—have gained increasing attention. This interest is likely to heighten in the coming years as NCLB’s “restructuring” provisions begin to take effect for thousands of schools. These provisions identify contracting with ESPs or converting to charter schools as two options that are available to improve chronically low-performing schools.

Whole school improvement models, such as the seven ESPs described in this report, are of great interest to decision makers because of the models’ claims that they are research-based and provide the training and other supports needed to encourage a coordinated approach to achieve student success. The research evidence indicates that some whole school improvement models are more effective than others and results vary greatly—even for effective models—depending on the quality of implementation (see Desimone, 2000).

### Conclusions

To date, education stakeholders at the national, state, and local levels have had few objective, rigorous, and consumer-friendly sources to turn to when making choices from among the ESP models that are available for adoption. This report serves as a consumer guide to help meet this need.

Sorting through and making sense of claims made by researchers and model developers is hard work, even for research scientists with years of training and experience. Despite substantial advances in developing standards and processes for judging and adding up the evidence in education, areas of disagreement exist in the research community. Even when the procedures necessary for reviewing and comparing large numbers of studies exist, the process is often complex and painstaking. Therefore, education decision makers often turn to others, such as the CSRQ Center, to sort through the evidence and report it as actionable information.

Readers should keep in mind the following points and limitations as they use this guide as a decision-making tool:

- **Only one ESP model (Edison) that was reviewed in this report had a solid body of evidence.** The evidence base for ESPs is emerging and the “jury is out” on the effectiveness of many of these models. All programs that claim to improve student achievement, including ESPs, will be increasingly challenged to demonstrate effectiveness based on rigorous studies.
- **The CSRQ Center’s standards were deliberately set very high.** We felt that in an environment that requires scientifically based research for many model adoption decisions, it was important to give consumers confidence that models in this report were reviewed against stringent standards.
- **The same high standards are used in all reports from the CSRQ Center.** Therefore, readers now have an opportunity to compare the ratings of ESPs with those already provided in our previous report on elementary school CSR models and with ratings in future reports on middle and high school improvement models—to be issued in fall 2006.
- **A low rating does not necessarily mean that a model is less effective than one with a high rating.** A low rating may mean that a model’s effectiveness has not yet been established through rigorous research. All ESPs should continue to conduct research to ensure consumers that (a) the model may work under some conditions and (b) the model does work under most conditions.

- **It was difficult to find publicly available information on many of the ESPs reviewed.** Several of the ESPs provided the CSRQ Center with information and insight that we believe will benefit education consumers. Other ESPs legitimately chose not to participate in our review process. As a result, we assigned a no rating in many dimensions of some ESPs because little publicly available information exists about the model.
- **The contracting process may provide a significant opportunity for education consumers to gather further comparative information on ESPs.** Education consumers that are interested in multiple ESPs can use the framework that we provide in our reviews to ask for additional information. For example, decision makers might consider asking for copies of internal evaluations, documents that identify links between the ESP's instructional design and a research base, detailed cost estimates, and other information that we were not able to provide.
- **The purpose in providing ratings in the CSRQ Center's reviews is to expand and clarify options for decision makers, not to dictate choices by picking winners and losers.** Models that received low ratings in one category may have high ratings in other categories. Education consumers should take a holistic view of the evidence presented across all five categories to make decisions about meeting locally defined needs.

The education community increasingly turns to research to help sort through its expanding school improvement options. This reliance on solid evidence helps to meet an NCLB provision that requires school improvement efforts to be driven by scientifically based research. However, and more importantly, using science to help support decision making will help to meet the urgently felt need on the part of educators and policymakers to ensure that their efforts improve the lives of children. The CSRQ Center issues consumer guides with the hope that the rigorous analysis and straightforward information that they contain will contribute to making research relevant to improving the education of our nation's students.

### Reference

Desimone, L. (2000). *Making comprehensive reform work* (Urban Diversity Series 112). New York: ERIC Clearinghouse on Urban Education. (ERIC Document Reproduction Service No. ED 441915)

**Table 1. Quantitative Study Findings Used to Rate Evidence of Overall Positive Effects on Student Achievement**

Education Service Provider	Number of Studies					Number of Findings	Percentage of Positive Findings
	Initially Relevant	Eligible for Full Review	Meeting Standards	Conclusive	Suggestive		
Edison Schools	28	17	9	7	2	19	40
Imagine Schools	0	0	0	0	0	N/A	N/A
The Leona Group, L.L.C.	2	1	0	0	0	N/A	N/A
Mosaica Schools	3	1	0	0	0	N/A	N/A
National Heritage Academies	3	1	0	0	0	N/A	N/A
SABIS Educational Systems, Inc.	0	0	0	0	0	N/A	N/A
White Hat Management (HOPE Academies)	1	0	0	0	0	N/A	N/A
<b>TOTAL</b>	<b>37</b>	<b>20</b>	<b>9</b>	<b>7</b>	<b>2</b>	<b>19</b>	

**Key:**

**Initially Relevant:** Of the nearly 940 studies screened, the number of studies per model found to be relevant to this review.

**Eligible for Full Review:** The number of studies per model that used research designs that were sufficiently rigorous and included student achievement outcomes.

**Meeting Standards:** The number of studies per model considered to be *suggestive* or *conclusive* according to causal validity rubrics of the CSRQ Center's Quality Review Tool.

**Conclusive:** The number of studies per model that used a rigorous research design (e.g., experimental, quasi-experimental) with no critical threats to validity.

**Suggestive:** The number of studies per model that used a less rigorous research design (e.g., longitudinal) with no critical threats to validity.

**Number of Findings:** The total number of individually measured outcomes found in the studies that met the CSRQ Center's standards.

The "N/A" indicates models in which zero studies met the CSRQ Center's standards.

**Percentage of Positive Findings:** The percentage of total findings in the studies that met the CSRQ Center's standards that were statistically significant and indicated that a model had a positive impact. The "N/A" indicates models in which zero studies met the CSRQ Center's standards.

**Table 2. Summary of Basic Information by Model**

Education Service Provider	Grade Levels Served	Number of Schools	Year Introduced in Schools	Costs (Year 1)	Evidence of Positive Overall Effects	Evidence of Positive Effects for Diverse Student Populations	Evidence of Positive Effects in Subject Areas	Evidence of Positive Effects on Additional Outcomes	Evidence of Positive Effects on Parent, Family, and Community Involvement	Evidence of Link Between Research and the Model's Design	Evidence of Readiness for Successful Implementation	Evidence of Professional Development/ Technical Assistance for Successful Implementation
Edison Schools	K-12	157	1995	Varies			Reading: Writing: Math:					
Imagine Schools	Pre-K-12	42	2004	N/A								
The Leona Group, L.L.C.	Pre-K-12	48	1996	Varies								
Mosaica Education	K-8	40	1997	Varies			Reading: Math:					
National Heritage Academies	K-8	51	1995	Varies								
SABIS Educational Systems, Inc.	Pre-K-12	31 <sup>2</sup>	1995	N/A								
White Hat Management (HOPE Academies)	K-11	32 <sup>3</sup>	1998	N/A								

**Note.** Readers are encouraged to use this table in conjunction with the entire report, which explains in detail how the approaches were reviewed and rated. The report also provides detailed information about each model's ratings and offers in-depth descriptions of each model's services.

<sup>1</sup>Although the rating in this subcategory is zero, readers should note that most of the studies on Edison that met standards and also demonstrated evidence of positive overall effects on student achievement, examined the effects of this model on schools that served primarily low-income and minority populations.

<sup>2</sup>SABIS operates 10 schools in the United States.

<sup>3</sup>White Hat currently operates 13 schools as HOPE Academies.

## Table 2. Summary of Basic Information by Model (continued)

**Key:**

**Grade Levels Served:** The grade levels served represents the full range of grades that the model serves.

**Number of Schools:** This reflects the number of schools that use the model, as reported by the model provider. This number includes all schools regardless of the length of time implemented or the level of implementation.

**Year Introduced in Schools:** This date refers to the year in which schools first implemented the model. This is included so that readers can judge whether the ratings are influenced by the relative newness of the model.

**Costs (Year 1):** The costs are estimates provided by the model provider. The full report provides additional details on costs for each model.

**Evidence of Positive Overall Effects:** This rating focuses on a model's overall effects on student achievement. The rating is a function of the number of studies that were rated to be *suggestive* and *conclusive*, the percentage of findings in the suggestive and conclusive studies that demonstrated a positive impact, and the average effect size of those findings. The final rating reflects the amount of rigorous research and the strength of the effects reported in that research. The full report provides complete information about the methodology used to produce all ratings in this report.

**Evidence of Positive Effects for Diverse Student Populations:** This rating refers to positive effects for the achievement of students from diverse backgrounds, such as low socioeconomic status, minority, special needs, or English language learners.

**Evidence of Positive Effects in Subject Areas:** This rating refers to positive effects on achievement in specific subject areas, such as reading, math, writing, science, or social studies.

**Evidence of Positive Effects on Additional Outcomes:** This rating refers to positive effects on additional outcomes, such as student discipline, student attendance, school climate, retention/promotion rates, or teacher satisfaction.

**Evidence of Positive Effects on Parent, Family, and Community Involvement:** This rating refers to positive effects for improvement in family and community involvement, such as involvement in school governance, participation in family nights, or homework support.

**Evidence of Link Between Research and the Model's Design:** This rating refers to evidence that the model developer can provide explicit links between research and the core components of the model. Core components are considered essential to successful implementation.

**Evidence of Readiness for Successful Implementation:** This rating refers to evidence that the model provider ensures initial commitment from schools, tracks and supports full implementation, and helps schools allocate resources for successful implementation.

**Evidence of Professional Development/Technical Assistance for Successful Implementation:** This rating refers to evidence that the model provider offers comprehensive training opportunities and supporting materials, ensures that professional development effectively supports full model implementation, and develops the school's internal capacity to provide professional development.

● = Very Strong   ● = Moderately Strong   ● = Moderate   ● = Limited   / = Zero   ⊖ = Negative   (NR) = No Rating