

# Methodology

THIS MATERIAL IS EXTRACTED FROM *CSRQ CENTER REPORT ON MIDDLE AND HIGH SCHOOL COMPREHENSIVE SCHOOL REFORM MODELS*

<http://www.csrq.org>

OCTOBER 2006



AMERICAN  
INSTITUTES  
FOR RESEARCH®

The contents of this report were developed by the Comprehensive School Reform Quality (CSRQ) Center, which is operated by the American Institutes for Research (AIR), under a grant from the U.S. Department of Education. The views expressed in this report do not necessarily reflect those of the U.S. Department of Education, AIR, or the CSRQ Center.

---

# Methodology

This report is the first systematic attempt to characterize the evidence on the effectiveness and quality of 18 widely used middle and high school whole-school reform models. Although this report is intended for a general readership, cutting edge scientific concepts and processes have been used to produce the reviews in this report. This section details the research methods that were used to support the reviews, highlights some of the challenges posed in conducting systematic reviews of evidence, and provides technical readers with the background needed to judge the quality of the scientific efforts.

Past systematic reviews of the effectiveness of comprehensive school reform (CSR) models have relied heavily on published and unpublished reports about specific CSR models. Most notably, work by Borman, Hewes, Overman, and Brown (2002) and Herman et al. (1999) compared the effectiveness of specific CSR models in raising student achievement. Building on such work, this report from the Comprehensive School Reform Quality (CSRQ) Center quantitatively evaluates CSR models and provides a qualitative narrative description of 18 CSR models.

The CSRQ Center's researchers recognize that student achievement is critical to education consumers. However, education consumers also rely on thorough descriptions of whole-school improvement models, such as those in this report, and they want to know how their schools may change if they implement a specific model. Educators also seek information about the experiences of other schools that implement whole-school improvement models. To meet this need, the CSRQ Center combined qualitative and quantitative research techniques to report on (a) the impact that middle and high school models have on student achievement and (b) the experiences of schools that implement these models. This approach

aligns with Creswell's five purposes for the use of multimethods (1994, p. 175):

- Triangulation, in seeking convergence of results
- Complementary, in that overlapping and different facets of a phenomenon may emerge
- Developmentally, wherein the first method is used sequentially to help inform the second method
- Initiation, wherein contradictions and fresh perspectives emerge
- Expansion, wherein the mixed methods add scope and breadth to a study

Through the use of multimethods, the CSRQ Center reviewed available evidence on middle and high school models to determine their effects on student achievement and to expand and fully describe the components of each model and the services that they offer to schools.

As described in the introduction, the CSRQ Center developed the Quality Review Tool (QRT), a three-part, multimethod tool to collect and analyze qualitative and quantitative data to evaluate the middle and high school models for the education consumer.

1. **QRT Part 1** is the *qualitative data collection* phase. The purpose of QRT Part 1 is to gather (a) supporting information on each middle and high school model from publicly available sources, model directors, and three districts or principals and (b) descriptive information about the middle and high school models, such as professional development, technical assistance, and research-based design.

2. **QRT Part 2** is the *quantitative data collection* phase. The purpose of QRT Part 2 is to conduct a systematic review of the literature on the effectiveness of a middle and high school model on student achievement, other outcomes—such as attendance and graduation rates—and parent, family, and community involvement.
3. **QRT Part 3** is the *data analysis* phase, in which the qualitative and quantitative data are synthesized to generate effectiveness ratings of the middle and high school model. These ratings (very strong, moderately strong, moderate, limited, zero, and no rating) are developed for several categories, including evidence of positive effects on student achievement, additional outcomes, and parent, family, and community outcomes; evidence of a link between research and the model's design; and evidence of the model's ability to provide services and support (e.g., readiness and professional development/technical assistance) to schools to enable successful implementation.

## **S**ample of Middle and High School Models

The CSRQ Center gathered an initial list of 37 middle and high school models by consulting previous reviews (Borman et al., 2002; Herman et al., 1999; Slavin & Fashola, 1998), the Northwest Regional Educational Laboratory's (NWREL) *Catalog of School Reform Models*; Southwest Educational Development Laboratory (SEDL); the National Forum to Accelerate Middle Grades Reform; and the Center for Education Reform. From this list, a final sample was selected by:

1. Exploring the replicability of the middle and high school model, as determined by the total number of states implementing the model

2. Determining market share, as defined by the total number of schools implementing the middle and high school model
3. Investigating the comprehensiveness of the middle and high school model's design

During each step of the information gathering process, researchers consulted previous reports, databases, and Web sites of the middle and high school models.

For step 1 (replicability), the CSRQ Center's researchers consulted information from the Web sites of each middle and high school model to determine whether the 37 middle and high school models from the initial list operated in three or more states. This step narrowed down the initial list from 37 to 33 middle school and high school models.

For step 2 (market share), the CSRQ Center's researchers searched the Web sites of each middle and high school model for information on the total number of schools that used the model. The CSRQ Center defined the selection criterion for market share as middle and high school models that operate in 40<sup>1</sup> or more schools. This step narrowed down the list from 33 to 26 middle and high school models.

For step 3 (comprehensiveness), the CSRQ Center's researchers examined whether the features of the middle and high school model's design met the following components identified by the U.S. Department of Education: governance, administrative services, technical assistance, classroom practices, professional development, leadership development, benchmarks/assessments, and curriculum (U.S. Department of Education, n.d.). For coding purposes, components were defined as follows:

- **Governance** was defined as operations and management conducted in schools. Key words associated

<sup>1</sup>One model, First Things First (FTF), operated in more than 40 schools when the CSRQ Center was selecting its sample. Since then, however, three FTF schools were destroyed by Hurricane Katrina.

with governance were operations, structure, management, scheduling, committees, blocks, and administration.

- **Technical assistance (TA)** was defined as classroom operational or management assistance through mentoring, coaching, or other services provided to teachers. Key words associated with TA were troubleshooting, coaching, and mentoring.
- **Classroom practices (CP)** was defined as pedagogical, structural, and behavioral management practices that a teacher enacts in a classroom. Key words associated with CP were pedagogy, classroom management, classroom structure, teaching strategies, and philosophy of instruction.
- **Professional development (PD)** was defined as teacher training on a specific topic. This training typically occurs in a workshop or conference environment. Key words associated with PD were training (on specific topics), conferences, and workshops.
- **Leadership development (LD)** was defined as administrative training or development for school personnel in leadership positions (principals, grade-level chairs, and lead teachers). Key words associated with LD were leadership training and/or development.
- **Benchmarks/assessments** was defined as tests and evaluations used to measure students' skills and understanding and academic progress. Key words associated with benchmarks/assessment were measurable goals, formative evaluation, and benchmarks of progress.
- **Curriculum** was defined as the scope and sequence of learning objectives and indicators, as well as materials provided for lessons to instruct such objectives. Key words associated with curriculum were materials, scope and sequence, standards, and learning objectives.

Each middle and high school model was given a point for each component or criterion that the model met based on information found on the model's Web site and additional resources, including, but not limited to, *An Educator's Guide to Schoolwide Reform* (Herman et al., 1999); *Show Me the Evidence! Proven and Promising Programs for America's Schools* (Slavin & Fashola, 1998); and Web sites of the U.S. Department of Education (<http://www.ed.gov>), NWREL (<http://www.nwrel.org>), and SEDL (<http://sedl.org>). Each middle and high school model that had five or more components in its design was included in the final sample. This step narrowed the list from 26 to 23 middle and high school models.

Upon further examination, the CSRQ Center withdrew five models from its final sample for the following reasons:

- Two of the models were interventions that targeted specific populations of students within a school.
- One of the models had recently been discontinued.
- One model had been purchased by another business.
- The CSRQ Center could not verify the number of middle and high schools implementing one model.

By eliminating these models, the sample for review was narrowed from 23 to 18 middle and high school models.

## RT Part 1: Qualitative Data Collection Phase

QRT Part 1 was the qualitative data collection phase. It included guidelines for (a) conversations with model directors and (b) the collection of artifacts from the models and additional information about the model from publicly available resources (Bogdan & Biklen, 1998; Creswell, 1994, 1998).

QRT Part 1, including the guidelines for phone conversations, conversation questions, and artifact lists, was pilot tested with a model provider that was reviewed in the *CSRQ Center Report on Elementary School Comprehensive School Reform Models* (<http://www.csrq.org/reports.asp>), which was released by the CSRQ Center in November 2005. Based on feedback from the pilot conversations, researchers at the CSRQ Center modified the qualitative data collection process. To develop a complete description of each middle and high school model in the sample, an experienced and trained qualitative researcher at the American Institutes for Research (AIR) provided training on information gathering techniques, coding artifacts, and synthesizing qualitative data. The qualitative researchers met weekly to ensure consistency across the qualitative data collection efforts.

For QRT Part 1, qualitative researchers performed four main steps:

1. **Complete an initial description of the middle and high school model by using a standardized form.** The CSRQ Center developed the Model Description Form, a comprehensive survey instrument for compiling existing information about a middle and high school model, including mission, history, market share, costs to the school, and design of each of the middle and high school model's components. For example, researchers gathered information about the middle and high school model's organization and governance, such as how the middle and high school model provides site-based autonomy, whether additional personnel are needed, and whether the middle and high school model requires changes to the structure of the school. For questions about professional development, researchers gathered information about which school personnel are required to attend professional development, what types of professional development are offered before and during implementation, and what strategies are available to help a school build capacity to provide its own professional development. In all, researchers gathered information about the middle and high school model's organization and governance; administrative services; professional development; technical assistance; curriculum; instruction; inclusion; technology; time and scheduling; instructional grouping; student assessment; data-based decision making; and parent, family, and community involvement. The researchers also requested benchmarks and explicit citations that link the model's design to a research base. The researchers completed this form using the Web sites of each middle and high school model and other publicly available information.
2. **Conduct a phone conversation with the provider of the middle and high school model to verify previously gathered information.** Conversations were structured around the Model Description Form (completed in step 1). On average, phone conversations lasted 90 minutes.
3. **Conduct phone conversations with three schools that use the middle and high school model.** The conversations verified information gathered in steps 1 and 2. The schools were randomly selected from a list provided by the middle and high school model. The conversations were guided by the Model Description Form.
4. **Complete a final description of the middle and high school model by using a standardized form.** The Model Description Form-Complete synthesized all sources of qualitative data, such as the conversations with the model provider and the three schools and the artifacts collected from the middle and high school model. The Model Description Form-Complete was checked for quality control twice to ensure that each item had 100% agreement between the two qualitative researchers. This form was then used to organize the data based on core components. According to the CSRQ Center's standards, core components are

considered essential to successful implementation of the model. Additionally, these data were coded to answer several questions:

- Is there a strong link between research and the middle and high school model's design?
- Does the middle and high school model track and support full implementation in all schools?
- Does the middle and high school model help schools allocate resources to implement the model?
- Does the middle and high school model provide comprehensive training opportunities and supporting materials?
- Does the middle and high school model develop the schools' internal capacity to provide professional development?

## **Q** RT Part 2: Quantitative Data Collection Phase

QRT Part 2 was the quantitative data collection phase. Using systematic review methods (Borman et al., 2002; Lipsey & Wilson, 2001), QRT Part 2 included protocols to conduct systematic literature reviews and to code research studies for statistical and causal validity information.

QRT Part 2, including the protocols for literature reviews and coding instruments, was pilot tested using the same whole-school improvement model provider as was used for the qualitative data collection efforts (QRT Part 1). Based on feedback from the pilot test and from the *CSRQ Center Report on Elementary Comprehensive School Reform Models*, the process for conducting the literature review was improved and the coding instruments were refined. An experienced and trained quantitative researcher at AIR conducted training on how to use the coding instruments to

ensure consistency in the data collection. The training included a presentation of the definitions of different research designs, causal validity issues, and background information on calculating effect size.

For QRT Part 2, quantitative researchers completed five main steps:

1. **Conduct a thorough literature search.** For each middle and high school model, quantitative researchers searched educational databases (e.g., JSTOR, ERIC, EBSCO, PsycInfo, SocioFile, NWREL, DAI) and Web-based repositories (e.g., Google, Yahoo, Google Scholar). From these sources, quantitative researchers screened for *initial relevance* nearly 1,500 article abstracts or summaries across the 18 models in the final sample. To pass the initial screen, the studies had to meet several criteria: be published or distributed between 1980 and September 2006, examine at least one of the middle and high school models being investigated, use quantitative methods, and be reported as a full-text research paper (i.e., not a PowerPoint presentation or executive summary). From these articles, researchers identified 198 articles to code. Of those, 197 were available and retrievable for coding. Appendix T provides a summary table of the number of articles that passed through each phase of the QRT Part 2 process.
2. **Complete a Study Description Outcome Form (SDOF), the first standardized coding sheet.** The CSRQ Center's quantitative researchers used the SDOF to code and document each study's research design, outcome variables, and demographic information. The CSRQ Center assigned a lead and secondary coder for each article. The SDOF was completed by the lead coder. Then, the secondary coder verified all the information for 100% agreement. At this stage of coding, the primary focus was to screen each study for a reliable research design. Studies that *were not eligible for full review* often were evaluations of implementation theories

supporting the middle and high school model with no quantitative data on outcomes or used research designs that were not sufficiently rigorous (e.g., one group pretest-posttest research designs). Research designs that passed this stage included experimental and quasi-experimental research designs with both pre- and posttests that evaluated the middle and high school model with a control group (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002) and longitudinal and cohort designs with multiple testing periods. Studies with research designs that passed this screen and included student achievement outcomes became eligible for full review. A total of 41 articles passed this step and were eligible for full coding in step 3.

3. **Complete the Quality Indicators Form (QLIF), the second standardized coding sheet.** Researchers used the QLIF to code studies that appeared to use rigorous research designs. The QLIF served two purposes: examine the quality of the research and gather statistical information. Researchers examined the quality of the research, such as the internal and external validity, face and psychometric validity of the outcome measures, and other quality indicators (Herman et al., 1999). Coders also collected statistical information, such as effect sizes reported by the authors or raw statistical information. For each study that was relevant for full review, two quantitative researchers independently coded one QLIF for each achievement outcome in that study.
4. **Reconcile the two QLIF coding sheets to attain 100% agreement on each coded item.** If the two quantitative researchers could not reach a consensus, a review coordinator reviewed the coding sheets to facilitate reconciliation. After the reconciliation process, a final QLIF reflected the 100% agreement.

5. **Rate each article on an overall causal validity score.** The final step was to systematically map the information from the final QLIF (the reconciled version) based on a set of rubrics designed to score each study for its causal validity (Shadish et al., 2002) as *conclusive*, *suggestive*, or *inconclusive*. Studies determined to be suggestive or conclusive met the CSRQ Center's standards for rigor of research design.

Conclusive studies had high levels of rigor; that is, experimental and quasi-experimental designs that had zero critical threats to validity and fewer than two non-critical threats to validity. Effect sizes were reported or calculated only from studies that had a conclusive causal validity rating (Cooper, 1998; Light & Pillemer, 1984; Shadish et al., 2002). If the researcher could not calculate an effect size because of missing data, then the researcher conducted one of the following steps: (a) contacted the author for the statistical information needed; (b) imputed missing data, particularly standard deviations and sample size, using protocols established in previous meta-analysis (Borman et al., 2002); or (c) chose not to include the study in the synthesis if options a and b were not feasible.

Suggestive studies were those that had zero critical threats but more than two noncritical threats. Studies without control groups, including longitudinal and cohort research designs, were capped at suggestive, unless the analytic techniques generated high levels of rigor.<sup>2</sup>

Inconclusive studies had critical threats to validity, such as using testing instruments with poor face validity and reliability, insufficient program fidelity, nonequivalence of treatment/control groups, lack of proper baseline, and/or timing of outcome measures that was less than 1 school year after middle and high school model implementation or less than 1 academic

<sup>2</sup>For example, backward-looking interrupted time series designs were considered more rigorous than longitudinal or longitudinal cohort studies that examined trends over time.

year between pretest and posttest. Noncritical threats to validity included historical events, disruption/novelty effects, instrumentation changes, maturation, selection bias, and statistical regression (Shadish et al., 2002).

## **Q** RT Part 3: Data Analysis Phase

QRT Part 3 synthesized the qualitative and quantitative data to evaluate each middle and high school model in five main categories.

1. Evidence of positive effects on student achievement:
  - a. Evidence of positive overall effects
  - b. Evidence of positive effects for diverse student populations
  - c. Evidence of positive effects for specific subject areas
2. Evidence of positive effects on additional outcomes (e.g., student discipline, student attendance, school climate, retention/promotion rates, and teacher satisfaction)
3. Evidence of positive effects on parent, family, and community involvement
4. Evidence of link between research and the model's design
5. Evidence of services and supports to schools to enable successful implementation:
  - a. Evidence of readiness for successful implementation
  - b. Evidence of professional development/technical assistance for successful implementation

Category 1 used the quantitative information gathered in QRT Part 2. For each middle and high school model in the sample, the quantitative information—including the number of studies coded, the number of

studies that were rated as suggestive or conclusive, the percentage of findings in the suggestive or conclusive studies that demonstrated a positive impact, and the average effect size of those significant findings—was mapped onto rubrics to determine what rating the model should receive—either very strong, moderately strong, moderate, limited, zero, or no rating—for effects on student achievement. Quantitative researchers systematically aggregated results according to the QRT 3 rubric for the overall effect by grade, subject (reading, writing, math, science, and social studies), and diverse student populations (e.g., high poverty, minority, learning disabled and other special needs, and urban and rural students).

Category 2 evaluated the positive effects of each middle and high school model on additional outcomes, and Category 3 evaluated the evidence of positive effects of each middle and high school model on parent, family, and community involvement. Similar to Category 1, quantitative researchers mapped onto rubrics the information about the number of studies that evaluated these outcome variables, the number of studies that were suggestive or conclusive, the percentage of findings that demonstrated a positive impact, and the average effect size of those positive findings.

In general, the rubrics for the quantitative information for Categories 1–3 were as follows:

- **Very Strong.** If a model had at least 10 studies that met the CSRQ Center's standards for rigor of research design, with at least 5 rated conclusive (and/or conclusive studies constituted at least 50% of the total studies coded), and 75% of the outcomes showed statistically significant positive model effects ( $p \leq .05$ ), with an overall mean model achievement effect size for positive effects that is greater than or equal to +0.25, then the model received a very strong rating. A very strong rating is symbolized by a fully shaded circle (●).

- **Moderately Strong.** If a model had 5 to 9 studies that met the CSRQ Center’s standards for rigor of research design, with at least 3 rated conclusive (and/or conclusive studies constituted at least 50% of the total studies coded), and 51% to 74% of the outcomes showed statistically significant positive model effects ( $p \leq .05$ ), with an overall mean model achievement effect size for positive effects that is between or equal to +0.20 and +0.24, then the model received a moderately strong rating. A moderately strong rating is symbolized by a three-fourths shaded circle (◐).
  - **Moderate.** If a model had 2 to 4 studies that met the CSRQ Center’s standards for rigor of research design, with at least 1 rated conclusive (and/or conclusive studies constituted at least 50% of the total studies coded), and 26% to 50% of the outcomes showed statistically significant positive model effects ( $p \leq .05$ ), with an overall mean model achievement effect size for positive effects that is between or equal to +0.15 and +0.19, then the model received a moderate rating. A moderate rating is symbolized by a half shaded circle (◑).
  - **Limited.** If a model had 1 study that met the CSRQ Center’s standards for rigor of research design and 1% to 25% of the outcomes showed statistically significant positive model effects ( $p \leq .05$ ), then the model received a limited rating. A limited rating is symbolized by a one-fourth shaded circle (◒).
  - **Zero.** If a model had zero studies that met the CSRQ Center’s standards for rigor of research design or 0% of the outcomes in the studies that met the CSRQ Center’s standards for rigor of research design showed statistically significant positive effects, as required for a limited rating, then the model received a zero rating. A zero rating is symbolized by a circle with a diagonal slash (⊘).
  - **Negative.** If a model had at least 10 studies that met the CSRQ Center’s standards for rigor of research design, with at least 5 rated conclusive (and/or conclusive studies constituted at least 50% of the total studies coded), and 75% of the outcomes showed statistically significant negative model effects ( $p \leq .05$ ), with an overall mean model achievement effect size of less than or equal to zero, then the model received a negative rating. A negative rating is symbolized by a circle with a minus sign (⊖). Studies that receive a negative rating suggest that the model has detrimental effects. In practice, this review did not find any evidence of this kind for any model.
  - **No Rating.** If a model had no studies (i.e., no evidence was available), then the model received a no rating. A no rating is symbolized by a circle with “NR” (⊙<sup>NR</sup>).
- Category 4 evaluates the link between research and the middle and high school model’s design. This category used the qualitative information from QRT Part 1. Qualitative researchers applied the information synthesized in the Model Description Form (from QRT Part 1) onto the following rubric.
- **Very Strong.** If a model provided documentation that explicitly described and convincingly supported links between the research base and *all* (100%) core components of its design, then the model received a very strong rating. A very strong rating is symbolized by a fully shaded circle (●).
  - **Moderately Strong.** If a model provided documentation that explicitly described and supported links between the research base and *most* (75%) of the core components of its design, then the model received a moderately strong rating. A moderately strong rating is symbolized by a three-fourths shaded circle (◐).

- **Moderate.** If a model provided documentation that explicitly described and supported links between the research base and *half* (50%) of the core components of its design, then the model received a moderate rating. A moderate rating is symbolized by a half shaded circle (◐).
- **Limited.** If a model provided documentation that explicitly described and supported links between the research base and *less than half* (below 50%) of the core components of its design, then it received a limited rating. A limited rating is symbolized by a one-fourth shaded circle (◑).
- **Zero.** If a model provided documentation that referred to a *nonspecific* research base to support the inclusion of the core components in its design, then the model received a zero rating. A zero rating is symbolized by a circle with a diagonal slash (⊘).
- **No Rating.** If the CSRQ Center was unable to conduct a conversation with the model provider or obtain complete information to verify evidence, then the model received a no rating. A no rating is symbolized by a circle with “NR” (⊘<sup>NR</sup>).

Two main questions guided the ratings for Category 5 (evidence that the model provider offers services and support to schools to ensure successful implementation). The first question—does the middle and high school model provide evidence of readiness for successful implementation—included the following subcategories:

- Provider tracks and supports full implementation in schools.
- Provider helps schools allocate resources that are needed to fully implement the middle and high school model.
- Provider ensures initial commitment from schools.

Qualitative researchers used the information synthesized in the Model Description Form (from QRT Part 1) to rate the three subcategories using a specific rubric. These three ratings were then averaged to determine the rating for evidence of readiness for successful implementation. In general, a model’s rating was based on evidence of the following: formal or informal benchmarks for all or some of its core components; a formal or informal process for allocating such school resources as materials, staffing, and time; and a formal or informal process to ensure initial understanding of the model and commitment from staff.

The second question—does the middle and high school model provide schools with professional development and technical assistance needed to help teachers implement the model—included the following subcategories:

- Provider offers comprehensive training opportunities and supporting materials.
- Provider ensures that professional development effectively supports full model implementation.
- Provider develops a school’s internal capacity to provide professional development.

Again, each subcategory received a rating. The three ratings were averaged to determine the rating for evidence of professional development and technical assistance for successful implementation. In general, a model’s rating was based on evidence of the following: a variety of training opportunities, supporting materials for professional development in all or some of its core components, and a formal or informal plan to help build a school’s capacity to provide professional development.

In addition to the ratings across these five categories, the qualitative data gathered in QRT Part 1, such as the phone conversations and artifacts, were synthesized into a narrative description of each middle and high school model. Each narrative includes in-depth

information about the middle and high school model's costs and descriptions of the following components: organization and governance; curriculum and instruction; scheduling and grouping; technology; monitoring of student progress; parent, family, and community involvement; professional development and technical assistance; and implementation expectations and benchmarks.

In all, qualitative and quantitative data were mapped to rate each middle and high school model on five main outcomes:

- Evidence of positive effects on student achievement
- Evidence of positive effects on additional outcomes
- Evidence of positive effects on parent, family, and community outcomes
- Evidence of link between research and the model's design
- Evidence of services and support to schools to enable successful implementation

The quantitative analysis provided a systematic literature review of the reported effects of student achievement and other outcome variables. The middle and high school models that could show relatively more literature consisting of evaluation studies were more likely to achieve higher ratings in Categories 1–3 (as long as results demonstrated positive impact). Through the qualitative analysis, newer middle and high school models and those that did not have a substantial number of evaluation reports could be evaluated on such dimensions as quality of professional development offered by the middle and high school model. Although previous research on student achievement offers important considerations, education consumers may also consider whether the middle and high school model's design is based on solid research and provides a strong commitment to support schools through professional development and technical assistance.

Providers of newer models may not have had sufficient time to conduct research on the effectiveness of their models, but they should be able to clearly demonstrate that their models *can work*, that is, that the model's design is based on solid evidence of *what works*. Hence, by using both qualitative and quantitative methods, the CSRQ Center strives to provide the education consumer with a thorough and systematic description of the effectiveness of each middle and high school model reviewed in this report.

By using qualitative and quantitative methods to evaluate the effectiveness of widely implemented middle and high school models, this study also strives to provide usable information to education consumers. In 2005, U.S. Education Secretary Margaret Spellings stated that the No Child Left Behind Act “rests on the common sense principles of accountability for results, data-based decision making, high expectations for all, and empowering change” (U.S. Department of Education, 2005).

Meeting these goals will require a significant expansion of information for education consumers about what works. This report is intended to act as a decision-support tool for educators who wish to find effective whole-school improvement approaches for meeting locally defined needs. This report helps to provide descriptive and evidence-based information on selected middle and high school models that may help educators make decisions—marking a significant change in the culture of the education system to meet the needs of educators, policymakers, community leaders, families, and most importantly, America's children.

## References

- Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods* (3rd ed.). Boston: Allyn and Bacon.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). *Comprehensive school reform*

- and student achievement: A meta-analysis.* Baltimore: Center for Research on the Education of Students Placed At Risk.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.
- Cooper, H. (1998). *Synthesizing research* (3rd. ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* Thousand Oaks, CA: Sage.
- Herman, R., Aladjem, D., McMahon, P., Masem, E., Mulligan, I., O'Malley, A. S., et al. (1999). *An educators' guide to schoolwide reform.* Arlington, VA: Educational Research Service.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* New York: Houghton Mifflin.
- Slavin, R. F., & Fashola, O. S. (1998). *Show me the evidence! Proven and promising programs for America's schools.* Thousand Oaks, CA: Corwin Press.
- U.S. Department of Education. (n.d.). *Comprehensive School Reform Program.* Retrieved November 8, 2005, from <http://www.ed.gov/programs/compreform/2pager.html>
- U.S. Department of Education. (2005, March 10). *Expanding the promise, continuing the progress.* Retrieved November 2, 2005, from <http://www.ed.gov/news/speeches/2005/03/03102005.html>