

CSR Connection

The CSR Connection is an occasional paper of the National Clearinghouse for Comprehensive School Reform

<i>Introduction.....</i>	<i>1</i>
<i>Using Research in the School Improvement Process.....</i>	<i>2</i>
<i>Criteria for Scientifically Based Research in NCLB.....</i>	<i>3</i>
<i>Evaluating the Type of Quality of Research.....</i>	<i>7</i>
<i>Conclusion.....</i>	<i>8</i>
<i>References.....</i>	<i>10</i>

Primer for Scientifically Based Research

By Monica Martinez

Introduction

The No Child Left Behind (NCLB) Act of 2001 requires educators to practice research-based decision-making and program coordination, especially in low-performing, high-poverty schools. As a result, for schools to receive federal funds, educators must identify and plan to use programs and practices that have been proven, through research, to “work.”

Educators are being asked to become knowledgeable consumers of educational research who can identify high-quality research, particularly “scientifically based research” or SBR. This primer introduces the various forms of research to help educators identify research-based solutions, aligned with the federal legislation, to improve their schools.

NCCSR
2121 K Street, NW, Suite 250
Washington, DC 20037-1801
AskNCCSR@goodschools.gwu.edu
www.goodschools.gwu.edu
Telephone 1 (877) 766-4CSR or 4277
Fax 1 (877) 308-4995

About NCCSR — A partnership of The George Washington University and the the Institute for Educational Leadership.
The National Clearinghouse for Comprehensive School Reform collects and disseminates information that builds the capacity of schools to raise the academic achievement of all students. Through its web site, reference and retrieval services, and publications, NCCSR is the central gateway to information on CSR. If you have documents on CSR that should be added to our database, please contact us for submission information.

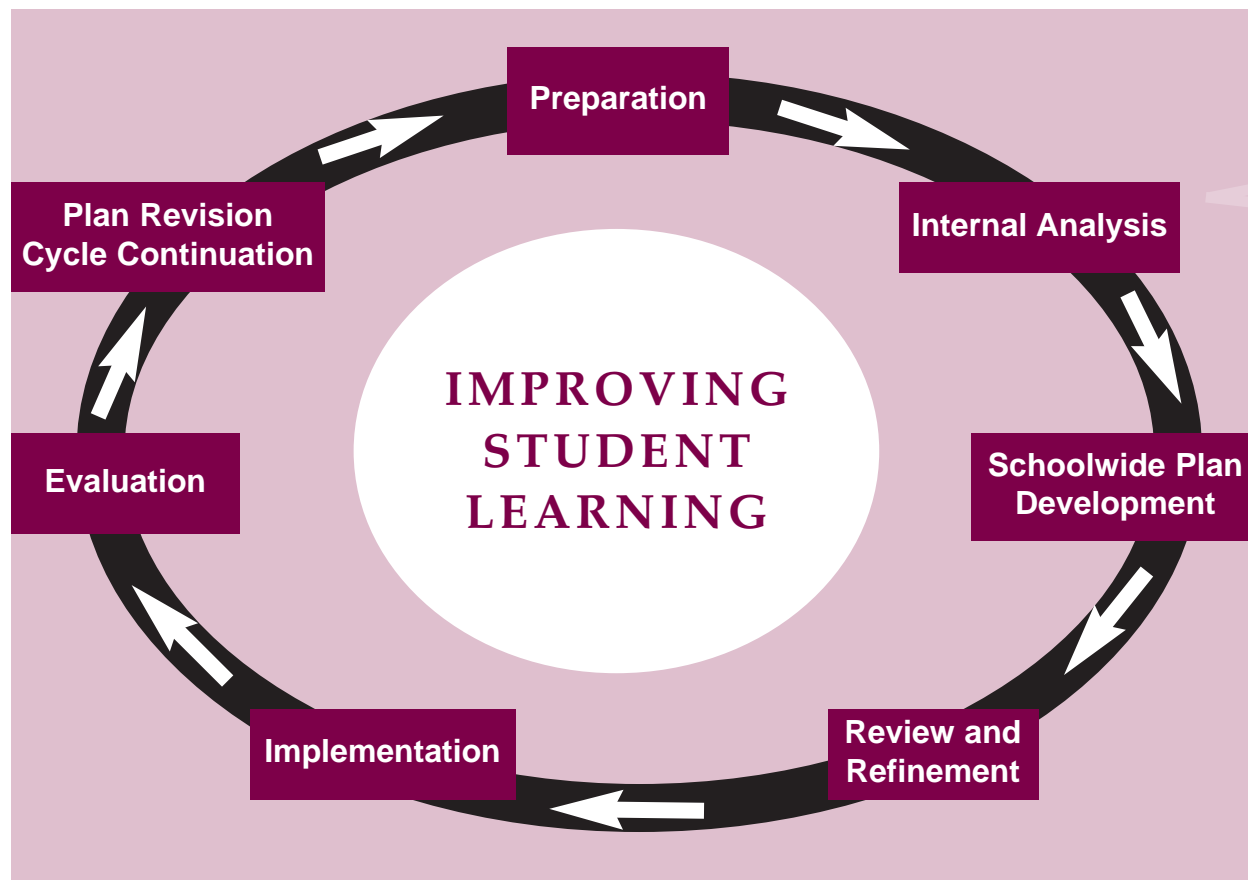


Figure 1. The Steps in Planning School Improvement

Using Research in the School Improvement Process

Research informs the school improvement planning process at various points and in different ways. Figure 1 illustrates the typical stages of school improvement planning and implementation—preparation, internal analysis, schoolwide plan development, review and refinement, implementation, evaluation, and plan revision. The process of school improvement is continuous, ongoing, and centered on the ultimate goal of improving student learning. Although research may play a role in each stage of the process, research-based decision-making will most typically influence the schoolwide plan development, review and refinement, and the evaluation stages.

A series of general questions guide school improvement planning and should be used to assess educational research:

- ***What are the current strengths and weaknesses of the school?***
- ***What types of programs and practices will improve student achievement?***
- ***How do these programs or practices work?***
- ***Where do these programs or practices work?***
- ***What type of population do they work for?***
- ***What resources are required to make them work?***

Using these questions to examine research can help educators identify whether or not a program or practice will work—have an

impact on student achievement for their specific student body. Factors to consider are the problem to be solved; the population demographics; the skill level, capacity, and experience of the teachers; the school context (urban, rural or suburban); and available school resources including district, state, federal, and private funds, facilities, technology, and books. The consideration of these questions is the starting point to examine scientifically based research.

Criteria for Scientifically Based Research in NCLB

Educational research is the formal, systematic application of the scientific method to the study of educational problems. Stated another way, educational research is the careful and diligent search for answers to questions about programs or practices that is based on evidence, not beliefs or opinions. While many types of research may be useful in making decisions regarding school improvement, NCLB established six specific criteria for the types of research that must be used if schools seek federal education funding, such as Comprehensive School Reform or Title I funding.

Importance of the Research Design

Educators should first examine the research design of any study. The research design criterion for scientifically based research is that it “uses experimental or quasi-experimental designs.” These research designs, which are types of quantitative research, attempt to ensure that measured differences in outcomes can be attributed to a particular practice or program. Experimental and quasi-experimental designs control the research conditions to purposefully limit the explanation of an outcome. The isolation of the intervention’s effect on an outcome is a key component of scientifically based research. For

example, a study conducted by Cook et al (1999) examined the theory of the Comer School Development Model that improvements in interpersonal relationships and social climate will lead to improved student achievement.

The 23 schools in the study were matched on the conditions of racial composition and student achievement scores for two consecutive years. The schools, all determined to have similar conditions, were then randomly assigned into the control group or the experimental group. The schools in the experimental group used the Comer School Development Model, while schools in the control group did not. The conditions in the Comer schools were controlled with regard to model implementation by an on-site facilitator who helped maintain fidelity to the model. The researchers used end-of-year teacher and student surveys in Comer (experimental) schools and non-Comer (control) schools as well as test scores to study the effect of the model on student achievement. Using this method, any changes in the school results could be attributed to the intervention, the Comer model, and the effectiveness of the program could be assessed.

Quantitative research, as in the Comer study, typically begins with a theory about a relationship that suggests certain data to collect. The research involves collecting numerical data to test a hypothesis or answer research questions about the current status of a subject or the intervention—the practice or program. In the Comer study, teacher and study survey data were collected, along with student test scores, to analyze the interpersonal relationships and social climates in the schools to test the theory that improvements in those areas would improve student achievement. In a quantitative study, the common methods used to test a theory are experiments using standardized instruments, such as achievement tests, surveys or highly structured interviews and observations. Once the data are collected and analyzed, the findings may either refute or support the theory. The data analysis is mainly statistical and

relies predominantly on a deductive approach, which derives specific conclusions from the research conducted.

Although only experimental and quasi-experimental research designs meet NCLB's criteria for SBR, a variety of research designs, quantitative and qualitative, can be used in conducting research for different purposes. For instance, descriptive studies assess the implementation or replicability of a practice or program and therefore must focus on issues of process and context. For example, qualitative studies can answer such questions as: How and in what setting was a program or practice implemented? How can an intervention be implemented or replicated in another school?

Qualitative research is usually multi-method in focus and involves an interpretive, naturalistic approach to its subject matter. Qualitative researchers study subjects in their natural settings, attempting to make sense of or interpret phenomena in terms of the meanings people bring to them. Qualitative inquiry often starts out with "how" or "what" questions; therefore, the focus is on making sense of and describing what is occurring. The qualitative approach is inductive as it moves from observations to more abstract generalizations and ideas.

In education, qualitative research is often used to understand how an intervention was implemented or interpreted. For example, a study by Wolf et al (2000) was conducted to understand how four schools successfully met the challenge of the Kentucky Education Reform Act (KERA) and to identify the factors or conditions that led to successful compliance. The study examined the four schools and teachers in their contexts to identify, understand, and derive explanations for the success of the implementation. The research method involved four case studies that required gathering data, through observation, on the relationships between teachers and students and teachers and school leadership; interviews; and the col-

lection of artifacts such as lesson plans and student work. In a case study, researchers "triangulate" the data by gathering data through multiple sources to gain confirming evidence for the factors that they are examining. To draw conclusions, researchers identify themes in the data and identify common patterns and notable differences to draw conclusions. In this study, researchers found factors common to the teachers in all four schools and also that each school had their own unique way of meeting the challenges of the reform initiative.

The NCLB legislation, however, requires the use of research with experimental or quasi-experimental designs to ensure that the practices and programs adopted by schools have demonstrable effects. By basing decisions on this type of research, it is presumed that educators can adopt an improvement strategy whereby they can predict, with greater confidence, whether the selected practice or program for implementation will work and if so, what the outcomes will likely be. The research design, therefore, should be the very first criterion examined.

Examining the Other NCLB Criteria

Only after it is verified that a study meets the research design criteria established under NCLB should the study be measured against the remaining five criteria:

- ***Has been through a rigorous peer review***
- ***Employs systematic methods***
- ***Is based on reliable and valid measurements***
- ***Uses rigorous analysis***
- ***Has a clearly defined methodology.***

Peer Review

The next criterion to consider is whether the research has been reviewed and accepted through peer review. This is, perhaps, the second most important criterion

of the SBR standards and the most efficient method to identify quality research. The purpose of peer review is to critique what researchers have done. In peer review processes, a study is sent to members of a review committee who provide feedback and recommendations. When an article is published in a peer-reviewed journal, the journal certifies the soundness of the research design and methodology.

This criterion suggests, therefore, that studies that are conducted internally and that do not undergo peer review should be regarded with more skepticism than those that are published in reputable peer-reviewed journals. In many cases, the people and organizations who have created a program or practice conduct an internal evaluation or commission another institution to conduct an evaluation. These evaluations may or may not have a rigorous research design, yet are published and cited as proof of the benefits of the practice or program. If a study or evaluation demonstrating the effects of a practice, program, curriculum or model has not been held up to outside scrutiny, this may serve as a red flag concerning the substance, generalizability and replicability of the findings as well the alignment to the criterion set forth by NCLB. Research results may also be presented and discussed at professional conferences, which may serve as another form of peer review, though potentially not as rigorous as the formalized review conducted by peer-reviewed journals. In recent years, much research is being published electronically. Not all electronic journals are peer reviewed, and consequently it is important to screen any research to ensure that it has been peer reviewed.

Systematic and Empirical Methods

The second criterion for scientifically based research is that it employs systematic and empirical methods. Systematic methods are consistent and professional ways of collecting and analyzing data. Empirical methods are grounded in numerical data, which can be used to verify or analyze observations. Where and

how data is collected is very important as it influences the conclusions and generalizations that can be drawn from the research study. Data collection issues include the composition of the research subjects or the sample of a study, how data were collected and measured, and the methods of analysis used to draw conclusions. For example, if a study aimed to provide descriptive data about fourth graders in an urban area, the methods should ensure that the sample of fourth graders chosen for the study are representative of the larger population of fourth graders so that conclusions can be drawn about the entire population.

For a study to meet this criterion, researchers must establish a clear and justifiable link between the research question and the method used to answer it. For example, if the purpose of the study was to assess the effectiveness of a program intended to raise student achievement, the research method should control for conditions in the experiment including the environment, assignment of subjects, and the assignment of subjects to two comparable groups. This method would allow the researcher to establish a control group and an experimental group to test the effectiveness of the program. In evaluating educational research, a consumer must determine whether the methods of collecting and analyzing data are aligned to the purpose of the study.

Reliable and Valid Measurements

Measurement is the process of turning concepts like student achievement into precisely defined variables. The third criterion for SBR is that the research “relies on measurements that provide reliable and valid data.” Reliability and validity attest to the integrity of the data. Reliability addresses the dependability or consistency of a measure used as part of a study. If a measure is reliable, it will yield the same result each time it is used. For instance, when a question is repeated in several different ways in a survey, the researchers are testing for consistency and reliability. It is important to remember that reliabili-

ty is about the measure itself, not the findings. Reliability means that the measure consistently measures what it is designed to measure.

Validity is the degree to which an indicator accurately measures what it is intended to measure. A variable is a valid measure of a subject if it is relevant or appropriate as a representation of that subject. Validity is also about the measure itself and does not directly refer to the strength of the conclusions in the study. For example, student achievement test data can be used as a measure of comprehension of subject matter. To determine the validity of the student achievement tests as a measure, a researcher would compare the test questions with the curriculum. Most measures are field tested to ensure that the measures are reliable and valid. In the field test, a pilot study is conducted on a sample of a large group to determine if the measures consistently and accurately measure what they were intended to measure.

Rigorous Data Analysis

According to NCLB, for research to be considered scientifically based, it must employ a fourth criterion, "rigorous data analysis." As expressed by the Comprehensive School Reform program office, "even the highest quality data are of little value unless analyzed thoughtfully and carefully." Thoughtful and careful analysis uses the most reasonable statistical tools to answer the research question and most often presents statistically significant results, results that did not occur by chance, and draws appropriate conclusions. Descriptive data is collected and analyzed through systematic and empirical processes, using appropriate statistical tools, such as percentages, to answer the research question. For example, a research study that investigates whether or not student achievement has increased or decreased over time will likely display results in terms of the percentage of students who perform at, above, or below the basic skill level for a period of years.

In more rigorous studies, such as quasi-experimental or experimental studies, statistical methods can be employed to authenticate that an effect or outcome did not occur by chance or is "statistically significant." A statistically significant result means that the researchers can conclude with some certainty that the practice or program under study had an impact and that there is good evidence for the theory or question being examined. It is important to note the difference between statistical significance and educational significance. Even if the results are statistically significant, educators must evaluate for themselves whether or not the results are useful in their educational contexts. Therefore, educators must determine if and how the results can be used in their settings.

Description of Methods

For research to be considered scientifically based, the fifth criterion requires that researchers provide a clear, detailed description of their method, including what measurements were used and how the data were collected and analyzed. The information provided must be detailed enough to make it possible for another researcher to conduct the same study and achieve the same results. If an independent researcher is able to reproduce the results, this validates the research. It is important that the results can be reproduced. Replicating the steps of the research but achieving different results will invalidate the findings.

By following each of the criteria to determine if a study meets the federal guidelines for SBR, educators will know that the results in the research can be directly attributed to the practice or program. Therefore, educators can be more confident about the conclusions that the research study draws and about the type of results that they would expect if their school or district were to employ the same practice or program.

Evaluating the Type and Quality of Research

To identify research that meets the definition and criteria set forth in NCLB, a hierarchy of research designs using the Olympic medal system was established: gold, silver, and bronze.

- **Gold standard research** is quantitative research conducted using an experimental design and is considered SBR.
- **Silver standard research** is quantitative research that is conducted using a quasi-experimental design and is considered to be “promising research” within the context of NCLB.
- **Bronze standard research** is quantitative or qualitative research incorporating a descriptive design or case study. This type of research is not considered to be scientifically based research or even promising research, and bronze standard research may not be used in justifying a reform strategy, practice or program under the NCLB guidelines. As “supporting research,” however, it can provide valuable information to assist in making a decision about the relevance of a potential reform option.

Gold standard research, which uses an experimental research design, tests hypotheses by observing variables, conditions or characteristics to show that changes in one variable are caused by changes in another variable. In this type of research, causality between the practice or program studied and the outcome is clearly established. If the study shows evidence of effectiveness, then the practice or program “works.” An experimental design requires that the researcher create two equal conditions for the purpose of studying the impact of a particular intervention. To accomplish this, a researcher must be able to exert control over all conditions of the experiment including the environment, random assignment of subjects, and the assignment of subjects to two comparable groups: control and experimental. Examples of conditions to

be controlled in an experimental study conducted in an educational setting include the materials used, teacher training or experience, the classroom environment, the time spent using a particular intervention (practice or program), the type of school, and characteristics of the student population.

Because researchers often cannot analyze data from every member of a population, it is more efficient to choose a sample of the population and gather data from that sample. Random selection means that each individual is chosen entirely by chance and every member of the population has an equal probability of being included in the sample. Researchers use random sampling because it enables them to select a large statistically representative sample. The issue of representativeness is important because by reducing the bias that might occur if subjects were selected for particular reasons, the researcher can draw generalizations about the larger population from the research results.

In an experimental study, subjects are randomly assigned to either a control or experimental group. The control group is the group that does not receive the treatment (program or practice) under investigation. The experimental group is the group that receives the treatment and, if the experiment is successful, will show any effects of that treatment at the end of the experiment. As an example, a study conducted by Fuchs et al (1999) examined the effects of classroom-based performance-assessment-driven instruction. Sixteen teachers were randomly assigned to performance assessment (PA) or non-PA conditions. The control group, the non-PA teachers, provided a baseline for comparison to the experimental group, the PA teachers.

To control for the conditions, the researchers provided the teachers in the experimental group with instruction and technical assistance. The experimental group (the PA teachers) attended an initial

workshop, administered three PAs over the course of several months, and, after each PA, met with colleagues to score PAs and share ideas for providing student feedback and instruction. The control group did not use PA-driven instruction. By using two comparable groups, the researchers were able to evaluate the impact of the treatment—classroom-based performance-assessment-driven instruction.

Quasi-experimental research is a form of experimental research with one important difference: the researcher cannot control at least one of the three conditions (environment, treatment, or assignment of subjects) of an experimental design. This form of experimental research design, therefore, is considered silver standard research. For example, when a researcher is attempting to study the impact of a reading program on student achievement in the first grade, the researcher would compare two different first grade classrooms—one in a school that is implementing the program and one in a school that is not. However, if the researcher cannot assign students to one school over another, she or he cannot control which students are in the experimental group and which students are in the control group. Therefore, the study would not have random assignment of subjects, which is one of the three requirements for experimental design. Similarly, if the researcher could not manipulate the treatment (program or practice) or control the environment, the study would be considered quasi-experimental. Characterized by the CSR Program Office as "reasonable quality studies," silver standard research may suggest evidence of effectiveness, but the research design does not meet the gold standard ideal.

Descriptive studies describe phenomena as they exist and answer questions about the current status of phenomena. Descriptive research typically takes raw data and summarizes it in a manageable form. Frequencies, percentages, and averages are the most common forms of statistical data for this type of study.

Although descriptive research designs can provide a wealth of information, a disadvantage is that they cannot accurately assess the effectiveness of educational programs—they only assess the perception of effectiveness. This research does not meet the requirements set out by the U.S. Department of Education for rigor. This type of research design, therefore, falls under the category of "bronze standard," research that does not meet the criteria of experimental or quasi-experimental research designs. Because descriptive studies can still provide valuable information that may be used in decision-making about the relevance of a potential reform option, educators must employ their professional judgment when using this research.

Conclusion


The thrust of the NCLB legislation is that applications for federal funding for Comprehensive School Reform and Title I must demonstrate that the proposed approaches are supported by evidence of their effectiveness—the best available evidence in every case. The evidence must be drawn from scientifically based research (where available), which is based on an experimental study that has been published after an independent review. The NCLB legislation requires the use of scientifically based research or gold standard research to prove that programs and practices "work."

The research reality, however, is that there is a clear need for educational research that meets the criteria for SBR. According to the U. S. Department of Education (2002), it is estimated that of the research conducted in the field, only about 5% currently meets the gold standard criteria and only another 25% meets the silver standard threshold. The remaining 70% of educational research falls under the bronze standard category.

Within this policy context, educators must become critical consumers of educational

research and exercise professional judgment in identifying research-based strategies that can increase student achievement. In particular, this means examining research to determine its relevance to a particular school context, digesting and synthesizing research findings across studies, and incorporating the evidence collected in the school improvement planning process. At the same time, quasi-experimental and experimental studies take a long time to conduct and will not be available soon enough to address the urgency schools feel right now, particularly given the extent of the need for proven school improvement strategies. Given the limited volume of research that meets the gold or silver standard, professional judgment exercised in concert with the "best empirical evidence" becomes the compass that helps to navigate the school improvement process.

Identifying research-based strategies that increase student achievement requires a new form of capacity of state departments of education as well as at the district and school level. Organizations, federal research centers, regional educational laboratories, and universities must provide help and support in developing this capacity. These same organizations and institutions must also begin to publish their results in publications that are accessible to practitioners and provide sufficient narrative that unpacks the complexity of the study to provide some form of authenticity to the research conducted. As practitioners demand high-quality research, the researcher community will likely respond by conducting more research that meets the high standards set by the federal government.



While identifying research-based strategies requires states, districts, and schools to expend scarce time and resources examining not just the results of research studies, but the quality of the methods and strategies in the studies, educators can consider this work an important investment. When educators adopt interventions and programs that are backed by

scientifically based research, they can be confident that—with the investment of human resources and careful implementation—their school reform strategies will lead to improvements in student achievement.

References

- Cook, T. D., Habib, F., Phillips, M., Settersten, R. A., Shagle, S. C. (1999). Comer's School Development Program in Prince George's County, Maryland: A Theory-based Evaluation. *American Educational Research Journal* 36(3), 543-597.
- Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Dutka, S., & Kataroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal*, 36(3), 609-646.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002). Retrieved June 4, 2004, from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- U.S. Department of Education. (2002). Comprehensive School Reform (CSR) guidance. Retrieved June 4, 2004, from <http://www.ed.gov/programs/compreform/guidance/index.html>
- U.S. Department of Education. (2002). Summary of Regional Symposia for CSR State Coordinators April - May 2002. Retrieved June 4, 2004, from <http://www.ed.gov/programs/compreform/symposiasumm.html>
- Wolf, S. A., Borko, H., Elliot, R. L., McIver, M. C. (2000). "That Dog Won't Hunt!": Exemplary School Change Efforts within the Kentucky Reform. *American Educational Research Journal* 37(2), 349-393.

The primer is based on a workshop that was developed through a collaborative partnership between Regional Educational Laboratories (RELs) and Comprehensive School Reform Coordinators from select State Departments of Education. NCCSR was primarily responsible for conceptualizing, designing, and producing the workshop. Monica Martinez was the primary author of the workshop and is the author of this primer.

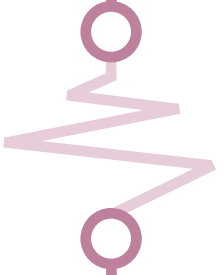
NCCSR collects and disseminates information that builds the capacity of schools to raise the academic achievement of all students. This is accomplished by continuously examining the literature related to comprehensive school reform (CSR), adding high quality materials to our on-line databases and actively sending useful information to educators and policy makers at the local, state and national levels. Through our web site, reference and retrieval services, and publications, NCCSR is the central gateway to information on CSR.

Contact NCCSR:
The National Clearinghouse for
Comprehensive School Reform
2121 K Street, NW, Suite 250
Washington, DC 20037-1801

Web Site:
<http://www.goodschools.gwu.edu>

Toll-Free Numbers:
Telephone: 1 (877) 766-4CSR
or 4277
Fax: 1 (877) 308-4995

CSR Connection is an occasional paper published by the National Clearinghouse for Comprehensive School Reform (NCCSR). NCCSR is funded by the U.S. Department of Education's Office of Elementary and Secondary Education and is operated by The George Washington University under Contract No. ED-99-CO-0137. The views expressed do not necessarily reflect the views of The George Washington University or the U.S. Department of Education. The mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government. Readers are free to duplicate and use these materials in keeping with accepted publication standards. NCCSR requests that proper credit be given in the event of reproduction.



National Clearinghouse for
Comprehensive School Reform
2121 K Street, NW, Suite 250
Washington, DC 20037-1801



Non-Profit Org.
US Postage
PAID
Wash. DC
Permit No. 593