CALIBER
ASSOCIATES

**DRAFT**

**EDUCATIONAL KNOWLEDGE BUILDING**

**IN CONTEXT**

Prepared by:

William A. Morill
Rhonda L. Munford

Caliber Associates
10530 Rosehaven Street
Suite 400
Fairfax, VA  22030
(703) 385-3200

**DRAFT**

Revised August 10, 2001

# TABLE OF CONTENTS

# EDUCATIONAL KNOWLEDGE BUILDING IN CONTEXT

## 1.      INTRODUCTION

This document is intended to set several important ongoing activities of the Department of Education (ED) in context, so that policy recommendations and decisions about each can be considered in light of their relationship to educational knowledge building goals and objectives. Specifically, these activities are the efforts of the National Educational Research Policy and Priorities Board (NERPP), the Office of Education Research and Improvement (OERI) and Planning and Evaluation Service (PES) within the Department to strengthen the quality and usefulness of national knowledge building that support improved educational experiences for all children. Cohesive and mutually supporting decision making about these related activities is important, particularly in light of the forthcoming congressional reauthorization of OERI.

In the broadest sense, there is a growing consensus that the quality, scope and depth of educational research and communication (referred to herein as "knowledge building") fails to meet the nation's needs for more widespread and higher levels of educational attainment to meet the requirements of the economy and modern life. In its reauthorization of OERI in 1994, Congress specifically mentioned this issue and mandated several systemic changes to improve knowledge building performance. The NERPP Board (also created by the 1994 Act) and other scholarly and distinguished external organizations have since called attention to the insufficient emphases on rigor, cumulative learning, effective communication and resource levels to pursue a knowledge building program. A variety of steps have been taken to attempt to strengthen the processes by which research is undertaken, assessed and integrated with educational practice. Other steps have been taken to sharpen the focus and critical mass of effort on urgent programs, and some efforts have been made to increase the resources to meet the perceived need more adequately, though with much more modest success in the case of OERI.

In the sections that follow, we call attention to particular efforts involving the NERPP Board's activities and responsibilities. They paint a picture of some progress, and identify the significant road ahead to meet identified goals. All are, in themselves, still works in progress, and together reflect a path toward a desired set of outcomes. We will first deal with effort to strengthen the peer review of the grants and contracts within OERI (known as the Phase I standards), followed by evaluation of the expert panel system to identify promising and exemplary practice (known as the Phase II standards). We then turn to an example of focused effort on a major priority area—comprehensive school reform (CSR)—to review work now underway. While only one of several areas of national focus (e.g., reading and math), CSR provides an interesting indicator of recent trends. Because the Board has not recently dealt with

this topic, we provide more detail of current OERI and Departmental activity. Finally, we turn to work now in progress at the Board's request on standards for research quality at the National Research Council, and close with summary observations.

## 2. STRENGTHENING THE PEER REVIEW STANDARDS

Title IX of the 1994 OERI reauthorization required that the Assistant Secretary, in collaboration with the NERPP Board, establish standards for evaluating applications for grants and cooperative agreements and proposals for contracts to undertake OERI work. Draft regulations were produced in June of 1995 and made final in October of 1995. These regulations set standards for peer review of grants, cooperative agreements and contracts, and made them applicable to field initiated studies (FIS), center grants and other programs. The standards specified requirements for peer reviewers, the processes by which they would operate, and the criteria to be applied.

After the completion of two rounds of competitions in FY 1996 and 1997 that encompassed both FIS applications and the centers, the NERPP Board sponsored a retrospective evaluation of the implementation of these standards. This 2-year review of a sample of the FIS applications (using 20 panels across five OERI institutes) and all of the center competitions involved substantial interviewing of reviewers, key Departmental officials and staff, applicants and other Federal officials and external organizations familiar with the processes. Findings and recommendations covered eight different areas:

- Enhancing the match between applications and reviewer expertise

- Reducing reviewer workload

- Bolstering professional development

- Clarifying the standards

- Modifying the review criteria and weightings

- Standardizing scores

- Providing feedback to unsuccessful applicants

- Exploring the use of technology.

The bulk of the findings and recommendations with one important exception identified practical and useful improvements to the established structure of the regulations, or suggested the

Department explore ideas that might also lead to similar improvements. They clustered around management improvements designed for operational improvement strengthening of the services needed by or useful to external reviewers.

One finding, however, went to a more fundamental matter in recommending the use of standing panels in each institute as a means of strengthening the expertise of the panels and reviewers in particular topic areas. This is also an approach used by NIH and others to strengthen the continuity of judgment from one year to the next. While this particular recommendation has not been adopted as originally stated, the staff can and does argue that it has adopted the spirit of the recommendation by establishing 11 areas of research activity for which a core group of panelists have been recruited for continuing long-term service. This step achieves most of the advantages of the standing panel without the formalities required by the Federal Advisory Committee Act (FACA). With that step taken, the Department could proceed, if it so desires, with the formal establishment of standing panels.

## Summary Comments

While peer review had been used previously by the Department, the legislated requirement and its implementation has put in place a uniform and formal process widely used to help ensure quality and equity in the review of applications and proposals in Federal procurements. The process is maturing with experience, and is a tool that can be helpful as part of a panoply of measures to increase or ensure quality. It should be noted, however, that peer review of applications alone cannot be expected to produce a level of quality that does not exist, nor is it an absolute requisite to assure it. Other work for the NERPP Board has identified Federal research organizations that produce high levels of quality without an external peer review system at all and, conversely, external peer review systems that produce less quality than desired or introduce other problems. The OERI system can be seen as a step in the right direction without supposing it is a cure-all for perceived problems.

## 3.    THE EXPERT PANEL SYSTEM FOR IDENTIFYING PROMISING AND EXEMPLARY PROGRAMS

Title IX of the 1994 OERI reauthorization also mandated and defined the basic structure for the identification, designation and dissemination of promising and exemplary educational programs. This new system replaced a less formalized system within the national diffusion network (NDN), and was initiated by Congress rather than the Administration. The underlying program responds to a widely shared goal of providing the most effective programs for the education of all children, and the complementary goal that programs so recommended come with

a known level of effectiveness. The growing urgency of improving educational outcomes for all children, and the problems ascertaining what works for whom gives this mandate a challenging urgency. While the Congressionally inspired legislation contained some detailed specifications as well as a basic structure for program implementation, considerable decision making about details were left to the Department.

Draft regulations were first posed in June 1996, and after a substantial review and comment period, final regulations were issued, effective in December 1997, when two pilot panels were already underway—Gender Equity and Math and Science. Two additional panels were subsequently launched—Safe, Disciplined and Drug-free Schools and Educational Technology. The entire first round was couched in a developmental process, though the production of final products in the legislative formulation depended upon the existence of a clear consensus about the standards by which "program," "promising" and "exemplary" could be judged, complemented by a compelling evidentiary base to support such judgments. As it became clear in an effort to devise a common set of standards across a diverse set of topics that no such consensus existed, the Secretary delegated such definitions to individual panel determinations. Thus, each of the four panels established its own meaning for:

- "educational programs" to be considered

- the specific meaning to be given to the four broad criteria for selection—quality of program, usefulness to others, educational significance and evidence of effectiveness and success

- specific criteria, if any, for distinguishing between "promising" and "exemplary" and

- the particular processes by which the reviews would be conducted within a common framework.

As in the case of the Phase I standards, the NERPP Board commissioned an evaluation of the expert panel system late in 1999, when the four panels were still underway (and some final actions still are in progress at this writing). The evaluation has involved extensive review of the written record and significant interviewing of panels, reviewers, a blind sample of applicants, involved Departmental officials and staff, and selected external officials with responsibilities for programs that might be considered comparable in purpose or process (there appear to be no exact matches). The evaluation analyzed the implementation of the expert review system by examining expectations against actual experience in four areas:

- Goals and objectives

- Scope and criteria

- Processes

- Staffing.

The evaluation then moved to structure a series of basic options for the expert panel system, from abandoning the system altogether as delivering more costs than benefits (if the level and rigor of the research program does not substantially improve) to a series of increasingly more substantial improvements. One option also would be to outsource the conduct of the program to one or more external organizations.

The evaluation found that the first four panels were necessarily developmental in character rather than reflective of a mature operational system. With respect to goals and objectives, the panels delivered a carefully considered set of recommended designations, though not without some selective controversy. While panelists were satisfied with the work they produced, most felt the evidence available to support their judgments was of less than desirable quality, reflecting the underlying concerns about methodological rigor and the limited pool of evaluations from which candidates could be drawn. Further, the evaluation noted a relatively loose linkage between the topics selected in the first round and the Department's critical priorities that should be given more attention as the system matures.

With respect to scope and criteria, the panels came to regard the criteria related to the evidence of effectiveness and success as the most important among the four, and indicated in the survey that they should become even more important. Indicative of this priority, the process created a separate Impact Review Panel, made available to all other panels, to help assess the applications on this criterion category. The evaluation further noted that this strong and appropriate emphasis suggested separate review and criteria for the exemplary and promising categories of programs.

In its review of the processes used in the expert panel system, the evaluation raised an issue about the long-term appropriateness of an application process rather than a field-wide review of exemplary programs that would further distinguish the process for promising and exemplary. It called attention to the substantial variation among panels—some warranted by the diversity of the topics, but others troublesome for assuring inter-panel fairness and consistency of judgment. The evaluation also noted the limited efforts to communicate the results compared to original expectations, and the absence of policy concerning the continuity and frequency of

future rounds of panel reviews. With respect to staffing, the evaluation commended much of what was done, but called attention to the importance of assuring the selection of individuals with current experience in teaching the topic under review as well as strong methodological expertise in assessing the evidence. It also recommended substantial reviewer training to assure consistent treatment of programs under review.

Overall, the evaluation recommended the continuation of the program as an important component of Federal knowledge building responsibilities and a response to widespread demand. It also recommended retention of Federal management of the effort, at least until several non-governmental initiatives to supply similar services mature to the point that they represent viable alternatives. The evaluation recommended a range of improvements to the system alluded to above, including the strengthening of criteria, linkage to other components of the research program, and communication with interested parties. It also recommended a phased change to a non-application approach to comprehensive assessments of exemplary programs, the separation of criteria and review cycles for promising and exemplary candidates, and legislative clarification that the judgments behind the designations are those of the panels rather than of the Secretary of Education.

## Summary Comments

The evaluation brings substantial focus to the issue of research rigor and the absence of a large pool of exemplary candidates with unambiguous evidence of effectiveness. This condition is due, in important ways, to lack of sufficient funds to mount evaluation efforts with strong designs and to the related lack of cumulative longitudinal research strategies. As the evaluation notes, without more rigor and substantially more funds, there will be precious little to consider in the exemplary category. This evaluation makes clear the importance of, and necessary linkage to, other activities addressed in this paper if this program is to succeed.

## 4.     SPECIFIC KNOWLEDGE BUILDING EFFORTS¾COMPREHENSIVE SCHOOL REFORM (CSR)

Among the central priorities for the Department has been the support of the broad national effort to encourage effective comprehensive and standards based reform for grades K through 12 through Federal initiatives, complementing such other priorities as early reading development, special education and, more recently, math.

The comprehensive school reform effort over the last decade grew out of conceptual research and foundation supported demonstrations that called attention to the importance of

integrating and aligning c comprehensive vision, the curriculum, professional development and assessment for effective educational performance.  This departure from incrementalism also demanded a substantial management effort, accompanied by more intensive technical assistance and applied research efforts at the school district level.

The Federal government began its support of CSR through the fiscal 1998 appropriations process in two important ways.  The Congressional action (known as Obie/Porter) made available funds to states for use in sub-grants of at least $50,000 to Title I eligible school districts to assist in planning and implementation of CSR programs meeting specific criteria concerning each program and its supporting infrastructure.  In an important complementary action, Title I resources could be used "school wide" to facilitate the CSR program efforts.

In addition, the Department launched a substantial knowledge building effort.  This CSR work now includes:

- OERI's research grants

- OERI's model design and evaluation contracts for middle and secondary schools

- OERI's capacity building grants

- ED's longitudinal evaluation of the effectiveness of school interventions (LEESI) to assess the federal demonstration program in the context of Title I schoolwide projects.

Another broader evaluation of CSR will soon follow LEESI, designed to provide a larger and more representative sample of all CSR efforts and to better determine answers to other questions such as district effects.

## 4.1    OERI Comprehensive School Reform Research Grants

A hallmark characteristic of CSR is ongoing evaluation to improve successful outcomes through research-based, whole-school reform.  Thus, it becomes necessary for all CSR models to remain actively engaged in a cycle of scientific study and review.  OERI grants totaling $21 million were awarded in late September 2000 to six research organizations: American Institutes for Research; The Education Alliance at Brown; Policy Studies Associates, Inc.; RAND; Success for All Foundation; and the University of Arizona.  OERI requires these funds to support independent and rigorous evaluation on widely implemented models.  Among other things, grantees are gathering data on student outcomes, changes in CSR schools, and the relationship between school characteristics and reform progress.

## Models Studied

In this round of research grants, a select set of currently adopted whole-school models are being examined. As listed in Exhibit 1, 14 models will be investigated. Only one grantee, Policy Studies Associates, Inc., has not specified its models to be examined. Four models— Accelerated Schools, Co-NECT, Modern Red Schoolhouse, and Success for All—are being studied by more than one grantee.

| EXHIBIT 1 ||
|---|---|
| **Comprehensive School Reform Model (N=14)** | **Grantee(s)** |
| Accelerated Schools | American Institutes for Research; The Education Alliance at Brown; RAND; University of Arizona |
| Atlas | American Institutes for Research |
| Audrey Cohen | University of Arizona |
| Coalition of Essential Schools | University of Arizona |
| Co-NECT | American Institutes for Research; The Education Alliance at Brown |
| Core Knowledge | RAND |
| Direct Instruction | RAND |
| Expeditionary Learning/Outward Bound | American Institutes for Research |
| Modern Red Schoolhouse | American Institutes for Research; The Education Alliance at Brown |
| Rigby Professional Development | The Education Alliance at Brown |
| Roots & Wings | University of Arizona |
| Success for All | American Institutes for Research; The Education Alliance at Brown; RAND; Success for All Foundation; University of Arizona |
| Turning Points | American Institutes for Research |
| Urban Learning Centers | American Institutes for Research |

## Topics Investigated

Collectively, these investigations, which range from 3.75 to 5 years, attempt to build on and expand current knowledge about:

- Overall model effectiveness

- Model effectiveness by student type, including but not limited to, English language proficiency, achiever-status (e.g. high achiever, low, or average), race/ethnicity, etc.

- Model effectiveness as a function of selected model components

- General factors that support or impede implementation

- Effect of teacher leadership patterns on model implementation

- Factors that support CSR

- Developer provided technical assistance

- Implementation of discrete CSR dimensions as a function of shared understanding of stakeholders' roles; district-level policy and action support; and other factors that support or impede dimension(s) implementation
- Model impact on a wide range of variables, including but not limited to, school climate, curricula, teaching practices, student test scores and academic outcomes, student attendance, student behavior, changes in teaching, etc.

- Factors that determine why and how a model is selected

- Analysis of theoretical underpinnings across models

- Factors that impact capacity to sustain the reforms.

Many of the topics investigated consider model implementation in some form or fashion—suggesting developers' keen awareness of the contribution of various contextual issues to successful outcomes. It is also encouraging to see growing numbers of researchers asking questions in relation to broader student demographic categories such as language proficiency and achiever status, as well as the adoption of experimental designs. Given the methodologies and types of research question, findings from this group of studies should yield a convergence of perspectives from key reform stakeholders including district-level leadership, school leadership, school instruction, and students. The types of questions further reflect the projected cycle of whole school reform (i.e., selection→implementation→support and technical assistance→ effectiveness/impact →capacity to sustain).

**Methodologies and Instrumentation**

Nearly all of the newly funded studies combine both descriptive/qualitative (i.e., "how things are") and relational/quantitative (i.e., "how things are in relation to other things") strategies to address their research questions. Half of the grantees describe efforts to use matched comparisons as part of their quasi-experimental studies.

One study conducted by the Success for All Foundation ("the Foundation") attempts to do something rarely carried out in educational research—impose randomization and collect pre- and post-test data. By using 20 experimental and 20 control schools clustered in the same districts (yielding approximately 4,200 student participants in each condition), the Foundation seeks unambiguous answers to the question "How did things get to be the way they are?"

In terms of sample characteristics, grantees are taking on school samples as large as 1200 (i.e., 600 CSR and 600 non-CSR sites) to as few as 40 (i.e., 20 newly awarded CSR schools and 20 non-funded schools). Some discuss efforts to infuse variation, yet balance, in terms of study sites (e.g., years of implementation) and student characteristics (e.g., socioeconomic status).

Finally, investigators describe a wide range of data gathering tools and instruments related to their studies. They include the following:

- Survey instruments and questionnaires

- Interview protocols

- Record reviews (including documents related to reform efforts, student achievement, school demographics, attendance records, and referral records)

- Classroom observations

- Focus groups

- Shadowing

- Site visits

- Pre- and post-tests (language & literacy and reading assessments).

## 4.2    OERI Model Design & Evaluation Contracts¾Middle and Secondary Schools

The prevailing collection of CSR models focus on improving outcomes at the elementary school level. Of the hundreds of existing national models, few target middle and secondary students. OERI is addressing this paucity by funding the development of new, research-based CSR models designed to meet the educational needs of students in middle and high school grades. Specifically, in 1999, OERI awarded seven contracts to develop and evaluate these models over five years, with a total budget of $76 million. Three of the seven models have been developed exclusively for middle schools. The remaining four designs have been planned for both middle and secondary schools.

The designs have similar features that reflect what is known about the learning and developmental needs of adolescent students. In particular, almost every model mentions as part of its essential components some relationship-building feature (e.g., smaller learning environments; teaching teams that follow students over time; long-standing, respectful relationships; safe and healthy learning climates). Each design places importance on high-

quality curricula, instruction, assessment, and professional development. Attention is also given to adaptability, flexibility, analysis, and refinement of the models.

Populations served by the models are identified as at-risk, urban, rural, Title I, and minority. None of the information provided about the models describes special student populations such as those with disabilities, limited English proficiency, or learning differences. It should be noted that, based on descriptions of populations served, primary objectives, and essential model components, the new designs seem to be paying closer attention to a persisting issue in the field—fostering greater achievement outcomes based on challenges peculiar to rural schools.

The awarded contracts are about halfway through the second year of funding. As such, there are no evaluation data available. However, contractors were required to describe planned evaluation activities, summarized in the next two tables. Exhibit 2 lists the focus of evaluation events and sets forth the number of contractors that plan to carry out a particular event. Exhibit 3 captures the instrumentation to be used, coupled with data on the number of contractors using each technique.

Nearly every model focuses on assessing the implementation of its design. Current evaluation plans emphasize measuring student achievement and overall impact of the model. All planned evaluations combine quantitative and qualitative techniques.

| EXHIBIT 2 | |
|---|---|
| **Focus of Evaluation** | **# Using (N=7)** |
| Design implementation | 6 |
| Implementation Support | 1 |
| Model's impact on culture/school outcomes | 2 |
| Model's impact on teaching practice | 2 |
| Model's impact on student behavior | 1 |
| Model's impact on student achievement | 4 |
| Development of small learning communities | 1 |
| Variation in scaling up the model across sites/planning and capacity building in scaling up | 2 |
| Overall effects/impact of model | 4 |
| Quality of model | 1 |

| EXHIBIT 3 | |
|---|---|
| **Instrumentation/Data Collection Methodology** | **# Using (N=7)** |
| Site Visits/Site Visit Interviews | 3 |
| Surveys | 6 |
| Interviews | 6 |
| Document Review/School Records/Historical Data | 4 |
| Student Achievement Data Records | 5 |
| Classroom Observations | 2 |
| Case Studies | 1 |
| Focus Groups | 1 |
| Policy Documents | 1 |
| Pre-Test/Post-Test Student Test Data | 1 |
| Professional Development Observations | 1 |
| Student Attendance Records | 1 |

## 4.3    OERI Comprehensive School Reform (CSR) Capacity Building Grants

All complex initiatives face serious problems in "scaling up." OERI defines this as capacity building—"improving the quality of reform models" (high quality) and "satisfying increased demand to adopt models" (high quantity). It has awarded $37 million in grants to 15 nationally recognized organizations. When applying for the grants, applicants were required to demonstrate that their models were "operating successfully in at least 15 schools with significant unmet demands for the applicant's services." Each of the awards spans a 3-year period that began September 30, 2000. Additionally, each grantee's project includes an evaluation component conducted by a third party.

**Efforts to Increase Model Scale**

Grantees are encouraged to pay attention to the following model features and processes when carrying out these efforts: accommodating more clients, ensuring tracking data, offering feedback on professional development, gauging the usefulness of materials (including the effectiveness for special populations and low-performing schools), providing staff support, and assuring evenness in the implementation over multiple sites.

Exhibit 4 highlights the grantees' proposed scaling up efforts by recording the current number of sites, listing their scaling up targets, and describing the grade levels currently served.

| | EXHIBIT 4 | | |
|---|---|---|---|
| **CSR Model** | **Number of Sites as of Fall 2000** | **Scaling Up Target(s)** | **Current Grade Levels Served** |
| America's Choice | 91 schools w/ many in NYC; Rochester, NY, DC, HI, and Jacksonville, FL | Add 100-150 schools | K through 12 |
| Atlas Communities | Nearly 100 schools in 13 states | Add 80 schools | K through 12 |
| Child Development Project | 28 schools in 4 states | Add 35-50 schools & prepare to add another 35-50 | K through 8 |
| Community for Learning Model | 136 schools | Add 6 regional demonstration professional development centers | Pre-K through 12 |
| Co-NECT | 135 schools in 27 states & DC | Add 175 schools | K through 12 |
| Different Ways of Knowing | 197 schools nationwide | Assist 80 new partnerships & 39 current relationships with upgraded technical assistance | Middle Grades |
| Effective Schools | 15 schools in 11 states | Add 15 schools | Tribal Schools on Indian Reservations |
| Literacy Collaborative Model | More than 250 schools in 100 districts | Add 20 schools with expanded grades 3 to 6 component | K through 2 |
| Modern Red Schoolhouse | 90 schools in 24 states | Add 34 sites | K through 12 |
| Onward to Excellence II | 26 schools in 12 states | Add 18 sites | K through 12 |
| Quest-Quality Educational Systems | 45 schools in 3 states | Add 14 rural middle schools in 14 States | Middle Schools |
| School Development Program | 400 schools in 21 states | Add 10 new rural school districts & 3 trainers and launch 3 Training & Learning Academies for 30 school districts | K through 12 |
| Success for All/Roots & Wings | 1,500 schools in all 50 states | No new sites targeted through this grant | K through 6 |
| Talent Development | 47 schools | Add 12 to 16 sites each year in large districts of several States including TX & NJ | 7 through 12 |
| Whole School Change | 122 schools | An additional 8 schools to encompass entire Boston school district | K through 12 |

Presently, the models serve as many as 1,500 schools in all 50 States (Success for All/Roots & Wings) to as few as 15 schools on Indian reservations in 11 States (Effective Schools). Additionally, scaling up targets range from as high as the addition of 175 schools over three years (Co-NECT) to as low as 8 schools in a single district (Whole School Change in the Boston school district). One grantee's scaling up efforts focused on the addition of six professional development demonstration centers (Communities for Learning). Another (School Development Program) is emphasizing increasing the number of districts in rural areas.

In addition to building new partnerships, Different Ways of Knowing is scaling up by increasing technical assistance in 39 of its current sites. Also, one grantee (Literary Collaborative) is adding 20 new sites with a newly developed intermediate program—grades 3 through 6, an expansion from the current K through 2 framework. Finally, it seems that, given its high level of dispersion throughout the nation, Success for All/Roots & Wings has opted to focus on improving the model's materials and training, bearing in mind State accountability measures, rather than increasing numbers of sites. Outreach efforts by model providers based on geographical need and geographical concentration were not discussed in the materials provided for this review.

## Efforts to Improve Model Quality

Grantees have proposed a wide range of activities geared toward improving their model's quality. They also reported on a primary objective to improve quality. From a review of the project abstracts, eight primary objective categories emerged. Exhibit IV-5 lists those eight areas of quality improvement, along with the associated model(s).

At least three or more grantees focused their primary objectives on: strengthening/ expanding specific model component(s); responding to diverse/special populations; developing/improving progress tracking and continuous improvement mechanisms; and refining the model's implementation process.

Some of the specific project activities mentioned to support the grantees' primary objectives included:

■ Developing Leadership Academies

■ Developing diagnostic tools to determine school readiness for reform

■ Creating data management resources

■ Providing a teaching/instructional clearinghouse of information

■ Establishing professional development demonstration sites

■ Exchanging online information

■ Developing video-based best practices information

- Revising and upgrading materials and curricula

- Increasing the ability to provide Web-based technical assistance and training.

| EXHIBIT 5 | |
|---|---|
| **Primary Objective Categories** | **Associated CSR Model** |
| Ensure models' staff gets needed orientation/training | America's Choice |
| Create/Improve Progress Tracking and/or Continuous Improvement Mechanisms | Atlas Communities<br>Different Ways of Knowing<br>Effective Schools<br>Onward to Excellence II |
| Refine Implementation Process | Atlas Communities<br>Co-NECT<br>School Development Program |
| Strengthen/Expand Specific Model Component(s) | Child Development Project (literacy component for ELLs)<br>Co-NECT (literacy & mathematics)<br>Effective Schools (C&I)<br>Literacy Collaborative (Expand to grades 3-6)<br>Quest-Quality (C&I/Alignment)<br>School Development Program (Instructional Services)<br>Success for All/Roots & Wings (New "K" program and upgrade other academic components)<br>Talent Development (smaller learning communities and improved partnerships)<br>Whole School Change |
| Improve Dissemination Approach | Child Development Project |
| Respond to Diverse/Special Student Populations | Community for Learning<br>Co-NECT (rural)<br>Effective Schools<br>Success for All/Roots & Wings (ELLs & SPED)<br>Talent Development (ELLs & SPED) |
| Enhance Technical Assistance/Professional Development | Community for Learning<br>Onward to Excellence II<br>Talent Development Model |
| Develop Web-Based Instructional Support | Modern Red Schoolhouse |

## 4.4 The Longitudinal Evaluation of the Effectiveness of School Interventions (LEESI)

In February 2001, the American Institutes for Research, RAND, and the National Opinion Research Center, (team for the longitudinal evaluation) submitted their Final Study Plan prepared under a contract with the Department. The project, known as LEESI, is aimed at contributing to a comprehensive evaluation of Title I schools. It involves an investigation of 100 Title I school-wide project schools with and without CSRD funding. It includes approximately 20 districts in five to seven States. With the exception of collecting pre-test data on kindergartners, data collection begins during spring of 2002, making the data gathering in sync with a large follow-on evaluation to be conducted by the Planning and Evaluation Service

mentioned earlier. It should be noted, however, that this study will follow all students enrolled as kindergartners (SY 2001-2002) in participating schools through the third grade (SY 2004-2005).

This paper reviews the proposed design of the LEESI study, paying particular attention to how it supports knowledge of comprehensive school reform. Conceptually, LEESI asks the following questions:

- What features of classroom practice affect the reading and mathematics achievement of students in Title I school-wide and CSRD schools?

- Are Title I and CSRD schools adopting strategies and interventions that improve classroom practice and student outcomes?

- How do State and district standards-based reform policies and practices required by Title I affect school strategies and interventions, classroom practices, and student outcomes?

**LEESI Design Overview**

This study focuses primarily on classroom practice and its relationship to student outcomes. The purposive sample of 100 schools will come from 20 to 30 districts in five to seven States. There will be five schools per district (three CSR/CSRD and two non-CSR/CSRD). The design will not account for whether the CSR is a national or locally/regionally designed program. As such, selected CSR schools will reflect a wide range of approaches within each district. Attention will be given to matching CSR schools with non-CSR counterparts in relation to demographics, etc.

At the State level, investigators will collect and review State policy records related to standards, curriculum frameworks, and assessments for the five to seven participating States. This will take place in the fall of 2002, 2003, and 2004.

At the district level, approximately 80 reform-related administrators/coordinators (e.g., 20 Title I coordinators, 20 reading coordinators, etc.) will participate in interviews in the spring of 2003, 2004, and 2005. Additionally, study investigators will collect a range of district level contextual documents for review and analysis during the spring of 2003, 2004, and 2005. Further, the design will limit variability in terms of States and districts to facilitate maximum variability at the school level. The study will only include States and districts that have made "substantial progress" in implementing components of standards-based reform.

School-level data efforts will include interviews with 100 principals, focus groups with six to eight parents from each of the participating schools, and collection and analysis of school-level aggregate scores on State- and district-wide assessments. Data collection will take place during spring 2003, 2004, and 2005. The sample size of 100 schools, however, will limit statistical power in determining school-level effects.

It is projected that the entire study will involve approximately 2,000 teachers, leading to substantial statistical power in determining effects on learning as a function of classroom practices. Among other things, teachers will complete surveys on school practices, on classroom practices in reading and mathematics, and on assessment of teachers' content and pedagogical knowledge. In total, there will be 13 data collection points using the full sample of teachers.

A subset of teachers might also participate in focus groups. In-depth observational tools will be used to ascertain information difficult to determine via survey instruments. Specifically, teacher logs, classroom observation, and videotaped and transcribed observations will be used. The sample of approximately 60 teachers each from grades 1 through 3 will participate in this level of data collection across eighteen data collection points.

To determine achievement outcomes, students will be administered reading and mathematics achievement tests by trained proctors. In addition to initial baseline assessment during kindergarten, students will be tested at six other data points (i.e., fall and spring semesters from fall 2002 to spring 2005). The study's research team is committed to investigating the use of a single reading and single math assessment across grades 1, 2, and 3. Moreover, efforts are underway to test Spanish-speaking, English language learners in the Spanish versions of the selected assessments. Finally, to the extent possible, this study will include the assessment of students with disabilities. The test administration conditions for this student group will match those of State and local assessments. In addition, during the spring of 2003, 2004, and 2005, schools will be asked to provide student record data on variables such as gender, ethnicity, attendance, grades, summer school and after-school participation, and eligibility for free or reduced price lunch. Student-level scores on state- or district-wide assessments will also be reviewed in the spring of 2003, 2004, and 2005.

## Summary Comments

The size and design characteristics of the collective work now underway in the Department on CSR provides an encouraging indicator of the response to the NERPP Board and other external urging to increase scope and rigor of the research and evaluation efforts of major national and Departmental priorities. On the other hand, there is still a great distance yet to be

traveled, as well as the associated time and resources needed to make the journey. This work is in its early stages with a number of years before producing conclusive findings. However, the scope and rigor of the designs in including longitudinal samples and comparison groups, careful documentation of implementation, and in-depth work to strengthen the interpretation of quantitative results are commendable and represent a serious effort to strengthen the quality of the evidence on critical educational issues.

Our review also makes clear how large the undertaking is, and how substantial the resources involved in strengthening the knowledge base about complex topics such as CSR. Some of the State plans express certain limitations in answering some important questions, and the collective output of all these States will need to be examined in order to grasp the full impact of the CSR undertakings. Future meta-analysis will be important to tie the results together. Project leaders are already hard-pressed to keep this substantial undertaking moving productively. It would appear useful in particular for the NERPP Board to urge careful data collection instrument construction on similar topics across projects to facilitate second round analysis, as hard as achieving that coordination may be.

## 5.     STANDARDS FOR RESEARCH QUALITY

The fourth component of the NERPP Board interest and activity is focused on the critical underlying issue of what constitutes appropriate research quality in educational research and how is it attained? To get a more comprehensive consideration of these questions, the NERPP Board and OERI commissioned a panel of the National Research Council (NRC) in the fall of 2000 to produce a report on this topic in late September or early October of 2001.

The NRC has also rapidly organized its work, and conducted a workshop March 7-8, 2001. The preface to the report of that workshop outlines the three questions to which the panel is addressing its work and deliberations:

- What are the principles of scientific quality in education research?

- How can research-based knowledge in education cumulate?

- How can a Federal research agency promote and protect scientific quality in the education research it supports?

The first question moves directly to the "what"—how do scientific norms, methods and traditions translate to education and what do they mean in the organization, synthesis and generalizability of education research? The second question addresses the critical issue of

cumulative learning, so critical to knowledge building in other fields.  The third, and perhaps the most critical of the questions centers on how OERI might proceed to develop and maintain the reputation and reality of sponsoring consistent quality and useful work.

**Summary Comments**

To the extent that the NRC panel produces a full and useful articulation of research standards and the ways to use them, those results can and should be integrated into the fiber of the three processes described above, thus helping to resolve troubling underlying issues in the peer review grant and contract selection process and the identification, designation and dissemination of promising and exemplary programs.  Such an articulation also becomes the basis for designing the research and evaluation agendas for major national knowledge building priorities and persuading funding authorities of the strong likelihood of high quality and useful results.

**6.      CONCLUDING OBSERVATIONS**

This analysis is intended to and, we believe, does, make clear the interconnections among the components of the NERPP Board activity and, more broadly, the knowledge building activities of the Department as a whole.  The success of any one of them is linked to attention and progress on the others.  There is every reason to be hopeful for improvement, if the steps reflected in this analysis are pursued, creating a momentum with cumulative impact on the entire system.  It is also important to recognize that the desired outcome cannot appear overnight and requires sustained effort and unflagging perseverance.  Such momentum will generate some clear-minded thinking not only of the end objective, but also the intermediate actions appropriate to a growing knowledge base.  What, for example, should we declare to be promising and exemplary in the meantime as we more carefully assess impact?  We believe that, central to the answers to this problem, is some unparalleled candor about the situation and the steps needed for significant improvement.