The Federal Mandate for Scientific Evidence

Sally M. Adkin, Ph.D., Sr. Vice President of External Programs

North Carolina School of Science and Mathematics, Durham, NC

adkin@ncssm.edu

A Paper Presentation for NECC, June 22, 2004, New Orleans, LA

Abstract

The enactment of the No Child Left Behind (NCLB) Act of 2001 and its central principal that federal funds should be used only for educational activities that can be supported by "scientifically based research" has focused extraordinary attention on educational research. The new federal mandates arising out of the NCLB legislation places quality research in high demand and presents challenging opportunities for educators. This paper describes the mandates and the characteristics of the more rigorous research methods and describes studies that are held up as models. Research and resources on learning technologies are listed in an annotated bibliography. Six research studies are cited to illustrate the key tenets of scientifically based research applied to the field of instructional technology. The author hopes this paper and the discussions that follow will result in practitioners being better informed and more hopeful that they can comply with the federal mandates and that better research studies will emerge to guide instructional technology.

Demand for Scientifically Based Research in Education

Educational research is under scrutiny. Our President and our nation have put forth the challenge to the educational community to revitalize the research base on what works/what is effective in educational practice. This paper will parse out the new federal mandates, provide guidance on understanding the criteria, and provide resources and examples for the march to comply. Examples will be drawn in particular from the field of educational technology. Evidence that the education community is to some extent attuned to the new force afoot is the assignment of an acronym for the phenomena. This paper will unabashedly use that acronym SBR to represent the term "scientifically based research."

Although the move to establish this more rigorous research paradigm has been underway for some time, the mandate itself is best embodied in the No Child Left Behind Act (NCLB, 2002) and the reauthorization and reorganization of the Office of Educational Research and Improvement (OERI) as the Institute for Education Science (IES)[1]. IES reflects the intent of the President and Congress to advance the field of education research, making it more rigorous in support of evidence-based education. The President appointed Grover J. (Russ) Whitehurst to a six-year term as the first IES Director.

In Senate testimony, Dr. Whitehurst (2002) argues that the current research base surrounding key educational issues is thin to nonexistent. What education leaders need is clear evidence of what works and what does not work, so that educators in decision-

[1] On November 5, 2002, President Bush signed into law the *Education Sciences Reform Act of 2002* establishing a new organization, the Institute of Education Sciences. The Office of Educational Research and Improvement, which had formerly been responsible for education research and statistics, expired upon enactment of the new Act.

making roles have clear guidance as they make curricular choices. Whitehurst points out that scholars in the field of reading have established a substantive and persuasive research base (NRP, 2002). Adherence to SBR will be a critical factor in the funding decisions and endorsement of new programs. NCLB refers to “scientifically-based research” 110 times with reference to Reading First programs for reading in grades K-3, Early Reading First for pre-K, Title I school improvement programs, and many more. In each case, schools, districts, and states must justify the programs they expect to implement under federal funding.

The federal mandate had two important tenets. See Table 1. First the literature base/ evidence for implementing new programs must be SBR, and secondly organizations who receive federal funds for their programs must demonstrate their effects through SBR.

Table 1

The Federal Mandate (NCLB, 2002) A. Interventions must be **based upon** findings from the best available scientifically-based research, and B. **Evidence** of the effectiveness of educational interventions must be demonstrated by using scientifically-based research methods.

The *based-upon* criterion is in itself a challenge for program implementers seeking trustworthy information. The *evidence-producing* criterion raises the bar to extraordinary levels and finds most program implementers deficient in needed SBR skills and confounded by important tenets of scientifically based research.

The Tenets of Scientifically Based Research

SBR is defined as "rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to the intervention," with an emphasis on evaluations that use experimental or quasi –experimental designs, preferably with random assignment (NCLB, 2002). Further SBR relies on measures that provide reliable and valid data and that offer documentation sufficient for review and replication. Questions posed should be based on identified needs and produce results of practical importance to educators.

The National Research Council's *Scientific Research in Education* (2002) provides a detailed report on what SBR means for educational research. A brief synopsis of key points from the report is set forth in Table 2.

Table 2

National Research Council. (2002) *Scientific Research in Education*
• Poses significant questions that can be investigated empirically • Links research to relevant theory • Uses methods that permit direct investigation • Provides a coherent and explicit chain of reasoning • Replicates and generalizes across studies • Encourages professional scrutiny and critique

The Council further notes that valid research for this purpose compares several schools that used a given program to several carefully matched control schools on meaningful measures of *achievement*.

The newly created IES imposes more rigorous standards in its checklist for establishing "strong" evidence. These are set forth in Table 3.

Table 3

Institute for Education Science. (2003). *User Friendly Guide—Identifying and implementing Educational Practices Supported by Rigorous Evidence*
• the outcome data are fully reported and valid • the size of the effect and the statistical significance are within appropriate ranges • implementation is at more than one site and these sites are typical school or community settings IES does allow as "possible" evidence: • very closely matched comparison-groups studies.

Resistance

These publications and other aligned federal technical assistance documents (Coalition for Evidence-Based Policy, 2002; CSR, 2002; NWREL, 2002; Raudenbush, 2002; Slavin, 2002) have provoked considerable criticism from the education research community (Jacob & White, 2002). The emphasis on SBR to shed light on student achievement holds up quantitative methods as the rigorous means of investigating causality. Simply put, many in the educational research community feel the SBR mandates are misguided attempts to constrain educational research (Jacob & White, 2002). The constraints include a constraint of outcomes to student achievement, a limit to causation studies, and the "gold-standard" of randomized assignment. Examination of the research models held up as those that meet the SBR standards suggest other limits. (See Appendix B.) The accepted measures of student achievement seem limited to validated standardized tests. Further, small studies are clearly of less interest than those with large sample pools and repeated trials.

A discussion of the arguments of the extended education research community is beyond the scope of this paper. Further, energies are better spent in understanding how to operate within the scope of the new mandates. Having devoted considerable time to reviewing the research in education and particularly in instructional technology, it is clear that there is a void in the kind of SBR causal research as described in this paper. That this is the only kind of research that can inform educators is clearly not true. However, that this is the only kind of research that is likely to attract much federal funding is true at this time. An investment in quantitative education research and the development of mechanisms to encourage evidence-based practices does hold promise in transformations in education of the same order and magnitude as those recently seen in medicine and agriculture and pointed to by IES Director Whitehurst (2002).

SBR Tools and Methods

The NCLB Legislation with its emphasis on SBR has resulted in several key governmental publications designed to provide educational practitioners guidance in distinguishing the kinds of studies that truly provide scientifically-based evidence. Many of these documents are listed in the reference section that follows this paper. Appendix A provides an important and brief primer of the tenets of SBR. This primer is aligned with the more rigorous standards set forth by IES. The primer is intended to serve as both a reference and check list for educational leaders. The key elements set forth in this appendix will be revisited in the discussion that follows of illustrative educational technology research studies.

Appendix B is an annotated bibliography describing four studies that are frequently referenced as conforming to the new federal SBR mandates. The documents in Appendix B are recommended for those educational leaders interested in studying more carefully what research looks like that conforms to the new standards and who has produced such research. The *What Works Clearing House* (2002) was newly established by the U.S. Department of Education to provide educators, policymakers, and the public with a central, independent, and trusted source of scientific evidence of what works in education. The clearing house is currently rather bare but holds promise to help educators make sound educational choices based on SBR.

Applying SBR to Inform Educational Technology

Technology in the form of computers entered America's classrooms at a time when accountability measures were an expected norm and when professional organizations were emerging to serve social scientists. Because of the relative expense of establishing technology in America's classrooms, the search for "effects" of the new technologies has received much attention. The proliferation of research has come up against the usual confounders in educational settings and is further complicated by the many kinds of technology tools, education purposes, and likely outcomes.

For example, there have been studies of computer-based software, including tutorial, drill and practice, and the more sophisticated integrated learning systems. All of these programs can individualize the education process to the levels and learning styles of the user. Other computer applications involve simulations and modeling. Some computer tools are highly creative, involving the student in design or team approaches. Probe ware for collecting and managing data has been widely adopted for science classes. Computer-based math tools such as graphing tools, spreadsheets, and calculators are quite different but important applications of instructional technology. Word processing, database tools, scanning technologies, graphics and photo software are among the many productivity tools that are finding their way into households and classrooms. Educators have developed many special classroom activities and collaborative projects based on the use of the Internet and the many resources available on the Web. Entire Web-based universities have now entered higher education and there is a proliferation of learning options supported by distance learning technologies. The North Central Regional Educational Laboratory (NCREL) has this to say about using technology to improve student achievement (NCREL, 1999):

> "Since educators first began to use computers in the classroom, researchers have tried to evaluate whether the use of educational technology has a significant and reliable impact on student achievement.

> Searching for an answer, researchers have realized that technology cannot be treated as a single independent variable, and that student achievement is gauged not only by how well students perform on standardized tests but also by students' ability to use higher-order thinking skills (such as thinking critically, analyzing, making inferences, and solving problems). Judging the impact of any particular technology requires an understanding of how it is used in the classroom and what learning goals are held by the educators involved, knowledge about the type of assessments that are used to evaluate improvements in student achievement, and an awareness of the complex nature of change in the school environment."

The search for effects of education technology is further challenged by the rapid growth of technologies appropriate for school settings. The escalation of school technology exceeds the current knowledge of how to effectively use or measure its impact. Technology evaluated a decade ago is different today than it was in the past, and educators can expect that when researchers take measure of current technologies the findings may not necessarily accurately reflect its future impact.

Many creditable groups have sprung up with a mission to explore and share information on emerging technologies and how technology can be used effectively to improve learning. An annotated bibliography of such organizations can be found in Appendix C. The bibliography describes the work and contributions of each group. Many of these organizations publish journals; all of them seek to support and inform educational uses of technology. Following the description of each organization, is a citation and link to an example of research that is located on the organization's web page.

For educational leaders and researchers, these resources are a creditable source for addressing the new federal mandate—both the **based upon** and **criteria for producing evidence** mandates. Works from these organizations would be appropriate to base school decisions about the adoption of educational technology interventions. The sites also provide models of how scientifically-based research can be produced. Table 4 is illustrative of some of the findings on effective integration of educational technology.

Table 4

Findings on Effective Use of Instructional Technologies

Teacher Reported Learning Outcomes Most Consistently Linked to Effective Use of Instructional Technologies (Adkin, 2000)
• Improved computer skills • Student's productions of sophisticated text and graphic documents • Finding sources and references • Students' broaden perspectives of the world • Engagement (vested) in classroom activities • Across-the-board academic gains

Further Findings on Effective use of Instructional Technologies from the Miliken Group (Schacter, 1999).
• Students learn more in less time when they receive computer-based instruction. • Instructional technologies that lead to the largest gains in student achievement are interactive technologies, learning experiences that require higher level reasoning, design tasks, and problem solving. • With the addition of instructional technology, teachers change their teaching practice toward more cooperative group work and less teacher stand-up lecture. • Consistent access to the technology, positive attitudes towards technology, and teacher training lead to the greatest student achievement gains. • The level of effectiveness of educational technology is influenced by the specific student population, the software design, the educator's role, and the level of student access to technology.

An Examination of Illustrative Educational Technology Research Studies

Having examined and provided a primer on the SBR federal mandate, the author will now turn to several illustrative studies in the field of educational technology. Six studies have been chosen for review. These studies are drawn from a literature search using the criteria: focuses on teaching and learning with technology, investigates the effects of technology on student achievement, implemented in a typical K-12 classroom, uses random assignment or strong comparative structure, has reported critical statistical data, and has been reviewed by creditable journal or organization. For a potential pool of such articles, the resources of the organizations listed in Appendix C were consulted.

Table 5 serves as an advanced organizer for these discussions. Each study is listed in the table followed by the student achievement content area, notes on the sample and sample size. N_T represents the number in the treatment group. N_C represents the number in the comparison group. The table has columns indicating how assignment was made to the treatment and comparison groups. Table 5 further sets forth notes on the measures used to gauge achievement, the precision of the findings, and comments on what groups have reviewed this study. This is a modest collection of studies and is not intended to represent the broad spectrum of instructional technology, nor necessarily the best studies available.

Table 5

Study	**Student Achieve Outcome**	**Sampling Frame & Appropriate Size**	**Randomly Assigned Comparison groups**	**Reliable/ Valid Measures**	**Precision**		**Review & Replication**
					Sig.	**Effect Size**	
Doty et. al.(2001) Interactive CD-ROM storybooks	Reading Comprehension	2^{nd} graders N_T=20 N_C=19 Size problems	Treatment Randomly assigned to class A or B	Stieglitz Informal Reading Inventory Test	Sig P<.05	.94	*Journal of Research on Computing in Education*
Koedinger, et al. (1997) Practical Alg. Tutor	Algebra I Achieve-ment	9^{th} graders N_T=470* N_C=120* *Less than half took final tests.	No, but 7 well matched comparison classes	Iowa Alg Aptitude Test & subset of Math SAT	Sig P<.05 Not Sig. P<.05	 .30 --	CARET & *International Jour. Of Art. Intell. In Ed.*
Kramarski, et. al.(2000) Reading in an Internet Environment	Eng. Reading Comprehension in Israel	8^{th} graders N_T=26 N_C=26	Treatment Randomly assigned to class A or B	12-item open test	Not Sig. P<.05	--	*Educational Media International*
McGraw, et. al.(2002) Interactive multimedia games on reading disorders	Reading & Writing Achieve-ment	6^{th} graders with dyslexia & ADHD N_T=~~15~~ 5 N_C=15 Size problems	Randomly Assigned	Process Assessment of Learner: Battery for Reading & Writing	Not Sig 21 out of 24 subtests p<.10	--	*Insight*
Scheidet, (2003) Global History & web-based environment	Global History Achieve-ment	10^{th} graders N_T=23 N_C=29 Size problems	Treatment Randomly assigned to class A or B	NY Regents Exam	Not Sig P<.05	--	*Journal of Research on Technology in Education*
Ysseldyke, et. al. (2003). Accelerated Math Software	Math Achieve-ment	4^{th} & 5^{th} graders N_T=157 N_C=61	No random assignment, but comparison groups	NW Achievement Levels Test STAR Math Exam	Sig P<.01	.40	*Journal of Educational Research*

Doty, D.E., Popplewell, S.R., & Byers, G.O. (2001). Interactive CD-ROM storybooks and young readers' reading comprehension. *Journal of Research on Computing in Education, 33,* 374-383.

The purpose of this study was to determine if there was a difference in the level of young readers' reading comprehension when the intervention group read an interactive CD-ROM storybook (*Living Books*) and a comparison group of students who read the same story from a conventionally printed book. The subjects were 39 second graders (two self contained classes) from a Title I elementary school in an urban school district in the US Midwest. Reading comprehension was measured through the use of oral retellings and six comprehension questions. A univariate analysis of variance was used to compare scores on oral retellings and comprehension questions. There was no significant difference in mean scores on the retellings between the two groups. There was a significant difference in mean scores on the comprehension questions between the two groups. When comprehension was measured through the use of comprehension questions; students reading the CD-ROM storybook scored higher. Evidence from this study indicates that the use of CD-Rom storybooks can have a positive effect on young readers.

Koedinger, K.R., Anderson, J.R., Hadley, W.H., & Mark, M.A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8,* 30-43. Available online at http://act.psy.cmu.edu/awpt/AlgebraPacket/kenPaper/paper.html

Koedinger examines the effectiveness of the Pittsburgh Urban Mathematics Project (PUMP), which incorporates the *Practical Algebra Tutor* (PAT) software program into the 9th grade algebra curriculum. PAT presents students with real-life situations, to which they are required to produce acceptable solutions and then provides the students with immediate feedback to their answers. The main elements of PUMP are an emphasis upon small group processes, the analysis of real-world problems, and the use of computational tools such as spreadsheets or graphing programs.

Koedinger compared PUMP student performance against seven comparisons classes; five "average" classes whose overall mathematics GPA from the previous year was roughly equal to the PUMP group, and two "exceptional" classes composed of students who had previously demonstrated exceptional academic achievement. All groups were roughly equivalent in terms of SES, and all seven comparison groups were taught in a "traditional" manner.

Based on standardized assessments (Iowa Algebra Aptitude Test and the Math SAT), PUMP students showed significantly higher mathematical ability than did the "average" groups. However, students in the "exceptional" group showed higher ability than did the PUMP group. Based on the Problem Situation Test, PUMP students' problem-solving ability was significantly greater than that of the "average" students, and equal to that of the "exceptional" students. Based on the Representations Test, PUMP students' ability to translate between equivalent algebraic representations was significantly greater than both the "average" and the "exceptional" comparison groups. It should be noted that the Problem Situation Test, and the Representations Test were both specifically designed to measure the mathematical skills that PAT was designed to promote.

Kramarski, B, & Feldman, Y. (2000). Internet in the classroom: Effects on reading comprehension, motivation, and metacognitive awareness. *Education Media International, 37,* 149-155.

The main goal of the research was to examine the contribution of an internet environment embedded with metacognitive instruction on students' reading comprehension, motivation and metacognitive awareness. The participants were 52 students who studied in two eighth-grade classes randomly selected from one junior high school and assigned to one of two conditions: (a) internet group (n=26) -exposed to metacognitive instruction embedded in an internet classroom; and (b) control group (n=26) - exposed to metacognitive instruction embedded in a regular class. Results indicate that although the Internet environment contributes significantly to the motivation of the students towards the study of English as a foreign language, no real contribution was found regarding actual improvement of achievement in the area of English reading comprehension and metacognitive awareness.

McGraw, T.; Burdette, K.; Seale, V.; Gregg, S. (2002). The effects of pervasive, consumer-based, interactive multimedia games on the reading disorders of ADHD children, *Insight, 2* (1), 7-30.

This exploratory study was conducted to determine if regular play of a popular, interactive multimedia game entitled *Dance Dance Revolution* (DDR) improved individual performance on reading and writing assessments of students with demonstrated reading impairment, dyslexia and Attention-Deficit Disorder (ADHD). The study population of 30 dyslexia + ADHD grade 6 students was drawn from a middle school in rural southwest Virginia. Fifteen (15) dyslexia/ADH students were randomly assigned to a treatment group and 15 dyslexia/ADH students to the control group. Intervention group data was included on only 5 students who completed all intervention sessions compared to the 15 in the control group. Both groups performed significantly better on 11 of 24 subtests based on the Process Assessment of the Learner: Test Battery for Reading and Writing (PAL-RW) test. The intervention group performed significantly better on only 3 of 24 subtests. Researchers reported no conclusions could be drawn and there was a need for further study.

Scheidet, R.A. (2003). Improving student achievement by infusion a web-based curriculum into global history. *Journal of Research on Technology in Education, 36,* 77-94.

This study investigated whether infusing Web-enhanced computer technology into the global history curriculum could increase the percentage of students achieving mastery level on the Global History NY Regents Exam and increase student interest and motivation in school work. The study focused on two sections of global history that were randomly and heterogeneously created by the guidance computer system. One of these sections was randomly selected as the experimental group (n=23) and was taught using a curriculum infused with Web-enabled computer technology, while the other global history section was taught in a traditional manner as a control group (n=29). Students in the experimental section utilized the Internet as the vehicle for doing independent research. Class time was devoted to students exploring Web sites and sharing discovered information in both large and small group settings for the purpose of solving authentic and challenging problems. Test results from the Global History Regents exam indicated a positive trend towards increasing the percentage of students achieving mastery level within a class totally infused with Web-enabled computer technology. However a t-test result of .69 when p=.05 and df =50, revealed no statistical significance between mean scores calculated for both the experimental and control groups. Practical significance is low based on an effect size of .20 between the experimental and control groups. Data from student and parent surveys, interviews, and classroom observations mutually supported the conclusion that there was a positive effect on student interest and motivations.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *The Journal of Educational Research, 96,* 163-173.

The authors examined the effects of implementing an instructional system that automates application of evidence-based components of effective instruction (*Accelerated Math* software) on student mathematics achievement and on classroom behaviors known to be related to overall student achievement outcomes. A treatment group of 157 4th- and 5th-grade students used the intervention in conjunction with the *Everyday Math* curriculum. Student performance was compared with a within-school control group of 61 4th- and 5th-grade students, as well as all 4th- and 5th-grade students in the district (N = 6,385). The students in the control groups received only the *Everyday Math* curriculum. Results indicate that the implementation of the instructional management system as an enhancement to Everyday Math resulted in an increase in the amount of time spent on classroom activities that researchers have identified as contributing to positive academic outcomes. Students who used the *Accelerated Math* program demonstrated greater mathematics achievement gains than did the control groups.

Discussion

Even empowered with what appeared to be a good pool of potential research studies drawn from credible organizations such as those listed in Appendix C, the author had to dig deeply to find a few studies that appeared to meet the study criteria. A considerable number of studies were rejected on first pass as not addressing student achievement or as not using quantitative methods. Over 200 such studies were passed over. This in itself demonstrates that the field of research in instructional technology is "thin" on the kind of SBR described in this paper.

Analysis of the modest collection of studies set forth in Table 5 and discussed above indicates that these studies do not hold up particularly well when measured by the SBR metrics set forth in the primer (Appendix A). A glaring weakness of all the studies is the small sample size. Only the Koedinger study of the Practical Algebra Tutor approaches the minimum sample size found in the primer of 150 students each per intervention and comparison groups. Note that Koedinger reports that less than half this number stayed with the study through final testing!

One strong element of these studies appeared to be that they did indeed use random assignment. When examined more closely, many used the valid technique of randomly assigning the treatment by classes not by individual students. Validly of this technique would require that every classroom had an equally likely chance to be selected for the intervention. That this design principal was followed is questionable. Random assignment by class, as noted in the primer, would require that 25-30 classrooms make up the study sample. This was clearly not the case in the studies under review. McGraw in the Interactive Games study judiciously applied the random assignment criteria, but then was only able to maintain a pool of 5 in the treatment group and 15 in the control group making the study worthless.

On reading carefully the author found that three out of the six studies produced results that were not significant. In other words an intervention that supposedly held great promise based on literature review failed to show an effect any greater than chance. Leaders in technology education are wide open to ridicule if 50% of the time research can not demonstrate that programs implemented make any difference in student achievement. Considering the number of studies that were passed over because the researchers did not even try to demonstrate a quantitative impact, it is very likely that the field of instructional technology research is demonstrating effectiveness no more than 50 % of the time.

On the bright side, these studies do provide some promising research design models and the implementers are at least trying to eek out quantitative effects. By finding the weaknesses in design and noting the discussions of possible flaws in treatment implementation, future researchers have a starting place.

A Challenge

Leaders in instructional technology need to disentangle themselves from the debate about the narrow and limiting confines imposed by the SBR federal mandates and step up to the plate to produce SBR that aligns with the newly imposed standards. A coordinated plan for research that intentionally launches studies that complement one another is important at this point. Much remains to be learned about the condition under which particular technology interventions have particular kinds of influences. The author hopes that this paper and the discussions that follow at NECC will result in practitioners being better informed and more hopeful that they can comply with the federal mandates. The presenter hopes the current bibliography will benefit by critique and will be enhanced through round table discussion with others who are interested in this topic and that more and better concrete examples will be forth coming.

References

Adkin, S. (2002). "Compelling and quantifiable effects of learning technologies". Unpublished paper presented at North Carolina Association for Educational Communications and Technologies (NCAECT) Leadership Evaluation Preconference, February 2002. Charlotte, NC.

Coalition for Evidence-Based Policy. (2002). Bringing *evidence-driven process to education: a recommended strategy for the U.S. Department of Education.* Author: Washington, DC.

CSR (Comprehensive School Reform Program Office), Office of Elementary and Secondary Education, U.S. Department of Education. (2002). *Scientifically based* *research and the comprehensive school reform (CSR) program.* Washington, D.C.: U.S. Department of Education. 18 pages. Available online at http://www.ed.gov/programs/compreform/guidance/appendc.pdf

IES (Institute for Education Science). (2003). *User friendly guide—Identifying and implementing educational practices supported by rigorous evidence.* Washington, DC: Coalition for Evidence-based Policy. Available online at www.excelgov.org/evidence

Jacob, E., and White S. (Eds.). (2002). Theme issue on scientific research in education. *Educational Researcher* 31, 8, p. 28. Available online at http://www.aera.net/pubs/er/toc/er3108.htm

NCREL (North Central Regional Educational Laboratory). (1999). Critical issue: Using technology to improve student achievement, *Pathways to School Improvement.* Posted 1999 on http://www.ncrel.org/sdrs/areas/issues/methods/technlgy/te800.htm

NRP (National Reading Panel). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development. Available online at http://www.nichd.nih.gov/publications/nrp/findings.htm

NRC (National Research Council). (2002). *Scientific research in education*. Committee on Scientific Principles for Education. Shavelson. R. and Towne, L. Eds. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press. Available online at http://www.nap.edu/books/0309082919/html/

NCLB-No Child Left Behind Act of 2001, Public Law 107-110, 115 Stat. 1425 (January 8, 2002) available online at http://www.ed.gov/legislation/ESEA02/

NWREL (Northwest Regional Laboratory). (2002). *Criteria for whole-school models.* Available online at http://www.nwrel.org/scpd/catalog/about/rubric2002.pdf

Raudenbush, S. (2002) "Scientifically-based research**.**" U.S. Department of Education Seminar on Scientifically-Based Research, February 2002. Washington, DC: U.S. Department of Education. Available online at http://www.ed.gov/print/nclb/methods/whatworks/research/page.html

Schacter, J. (1999). *The impact of education technology on student achievement-What the most current research has to say.* Saint Monica: Miliken Exchange on Education Technology. Available online at http://www.mff.org/pubs/ME161.pdf

Slavin, R.E. (2002). *Scientifically based research: A reader's guide*. Washington, DC: U.S. Department of Education.

Whitehurst, G. J. (2002). "Statement of Grover J. Whitehurst, Assistant Secretary for Research and Improvement, before the Senate Committee on Health, Education, Labor and Pensions." Washington, DC: U.S. Department of Education, June 2002. Available online at http://www.ed.gov/offices/IES/speeches

Appendices

Appendix A - Primer

Primer on the Tenets of Scientifically Based Research (SBR) and Conforming to the New SBR Federal Mandates

Appendix B- Some Models

Annotated Bibliography of Some Models of Scientifically Based Research (SBR) and Conforming to the New SBR Federal Mandates

Appendix C- Some Instructional Technology Research Resources

Annotated Bibliography of Leading Instructional Technology Research Groups and Representative Research from these Groups

Appendix A

Primer on the Tenets of Scientifically Based Research (SBR) and Conforming to the New SBR Federal Mandates

This primer is designed to help school leaders improve their understanding of SBR. The author has drawn from works of the Comprehensive School Reform Program, the Institute for Education Science, and the National Research Council. Complete citations are listed in the reference section of this paper.

Theory

Every scientific inquiry is linked to some overarching theory or conceptual framework that guides the entire investigation. Evidence should be presented that this theory is grounded in a solid understanding of methodological and empirical work that has come before. The theory should be set forth in clear terms along with the goals and activities central to the intervention and the contexts in which the intervention has been successfully implemented.

Design

The skepticism currently associated with educational research to a large extent focuses on questions of design. The first principal of design is that the design must fit the driving question. Not all good questions in education are causal and important contributions to understanding learners have been made through a variety of questions and quantitative and qualitative research methods designed to answer these questions.

The new federal SBR mandates call for questions to be of a causal nature. Study designs that are commonly used in educational research (of the causal nature)—such as pre-postal designs and most comparison-group designs—often produce erroneous results. (Coalition for Evidence-based Policy, 2002). Randomized controlled trials are the research design regarded by IES as the "gold standard" for assessing the effectiveness of interventions (IES, 2003). Issues of design are rooted in sampling frame.

Sampling

In any empirical research study, there is some set of individuals, institutions, or other entities that constitutes the object of investigation, e.g. 4th graders, known as the *target population*. There is also a definable group or aggregation of elements that are available to the researcher know as the *operational population*, e.g. 4th graders in my district. Populations are made up of elements termed *sampling units*. The sampling units in which a population is divided must be unique, in the sense that they do not overlap, and must, when aggregated, define the whole of the population of interest. Because of the way US public schools are organized, sampling units arise naturally in many educational research studies, e.g. the unique students from the list of those in our 4th grade classes. These are easily identified as unique and are readily counted. A list that uniquely identifies all the subjects in a finite population, in a particular order, is termed a *sampling frame*. The sampling frame is equivalent to the operational population and in fact defines the operational population.

Since it is rarely feasible to study an intervention with an entire population, some subset or sample must be drawn. The <u>size of the sample</u> needed to estimate a population effect with a specified level of precision is an important gauge as to the merit of the research. The theory of sampling allows a researcher to specify a sample size that is sufficiently large to ensure a high probability that errors of estimation can be kept within desired limits. A rough rule of thumb is that a sample size of at least 300 students (150 in the intervention group and 150 in the control group) is needed. If schools or classrooms, rather than individual students are randomized, a minimum sample size of 50-60 schools or classrooms (25-30 per intervention and control groups) is needed to obtain a valid finding. (IES, 2003). A defense of sample size should be included in any valid research study referencing the acceptable error limit, standard deviation of the data, and number in the operational pool. Even without a thorough understanding of the underlying calculations, the practitioner can make some reasonable judgments on whether the number of persons or units from which data is drawn seems reasonable to represent conclusions for the operational population.

Assignment of Comparison Groups

Among the tools and techniques deemed most valid for the elevated criteria for SBR is the careful use of *comparison groups*—a group that receives the intervention under study (the treatment group) and a group that does not (the comparison or control group). This requires splitting the operational population into two sampling frames that are identical and are both equivalent to the operational population. One technique commonly used to accomplish this *is random assignment* to each of the two groups such that every unit of the sample has an equal probability of being assigned to a given group. Random assignments go a long way to eliminate confounding variables. Researchers can never be sure that they have matched groups on some of the relevant *confounders*. Variables that predict getting the treatment that are also related to the outcome are confounders. In discussing the strength of random assignment, Randenbush (2002) points out random assignment applied judiciously eliminates confounders, all confounders even the ones not though of.

Simple random assignment is not likely to be the most efficient procedure available in an educational setting because of the way public schools are organized making it difficult to provide different interventions to students whose education is organized via classroom configurations. In a number of highly regarded randomized experimental studies, what has been randomly assigned to treatment will not be individual children, but will be schools. Although it may not be feasible to assign different 4th graders in any one school to different interventions, there are often promising programs that many schools want to implement but timing or resources will not allow for all interested to participate. So, during the interim period where one group of schools have started to do the program and the others are still waiting, the researcher has the potential for a randomized experiment, and a very ethically organized one (Randenbush, 2002).

Data Collection and Analysis

Even the strongest scientific research design is of little value unless *data collection and analysis* meet high standards. SBR relies on measurements or observational methods that provide reliable and valid data across evaluators and observers and across multiple measurements and observations. *Reliability* implies that repeated measurements on subjects taken under similar circumstances or over time will produce similar results. If unreliable, the data may hinder the researchers' ability to discern real differences among subjects or programs. To be considered *valid*, the tools and resulting data must measure the outcomes they were designed to measures, (e.g. that a 4th grader's math knowledge is what is being measured, not the ability to guess test answers.) There must be a match between the research question and the observed behavior on which the research findings are based. The data collected must be analyzed using methods that are appropriate for the task, adequate to test the stated hypothesis, and justify the general conclusions drawn. *Data analysis* methods must account for data complexities, for missing data, for changes in data over time, e.g., subjects that drop out of the study.

Precision

Before reformers make a final decision about the usefulness of available research findings, they must determine the level of precision of the study, i.e. the *significance.* Even high-quality research studies can produce findings that are not statistically or practically significant. Significance is a statistical term that helps readers to understand the likelihood that the findings of a study were the result of the designed intervention and would not be observed independent of that intervention. For practitioners, two standards of significance apply statistical significance and educational (practical) significance.

- For findings to be considered statistically significant, researchers hold to a high standard of significance at the .01 level, or an acceptable standard of significance at the .05 level. This means that the probability is less than 1% or 5% respectively that the observed difference might have happed by chance.

- For the finding to be considered educationally, or practically, significant, the effect on student achievement should be large enough to be of practical value. *Effect size*, calculated by comparing the mean of the control group and that of the treatment group, is a standard estimate of where the treatment group stands in comparison with the control group. For example, gains on a standardized test should be 10 percentile points to be considered educationally significant.

A study may have a significant statistical result in that the achievement gains reported in the study are very clearly the result of the reform program. However, the actual gains in achievement may be modest in size; when this is the case, the study may be said to have limited practical significance. In this context, school community members need to decide if the cost of implementing the program is outweighed by the size of the student gains that can be achieved.

Review and Replication

Even if a "strongly" scientific study demonstrates that the intervention had a significant impact, the researcher and reviewers need to develop alternative explanations for why the program worked. A search for essential implementation tenants, alternative explanations, and disconfirming evidence strengthen the findings. A strong study should be able to meet criticism by independent, *expert reviewers*. Peer reviewers, either from scientific journal or from an independent panel of experts in a given field, provide quality control in the form of a rigorous, objective, and scientific review of research.

Once an intervention has been shown effective, follow-up randomized trials are needed to establish the interventions' effectiveness as implemented on a larger scale, in typical settings. Only then can the education community be confident that the intervention will truly improve education outcomes. Scholars point out that the challenge of replicating interventions is generally greater in education and other areas of social science than in pure science. Social interventions, unfortunately, cannot be replicated quite as easily, and must therefore be rigorously evaluated for effectiveness when implemented on a large scale.

Appendix B

Annotated Bibliography of Some Models of Scientifically Based Research (SBR) and Conforming to the New SBR Federal Mandates

The following studies are among those that have been frequently cited in federal reports and technical assistance documents as examples of studies that model the new SBR federal mandates.

Tennessee Class Size Study (STAR)

Children were assigned at random to small classes (15 students), regular classes (20-25 students), or regular classes with an aide. STAR found that students in small classes did much better than students in regular classes in math and reading, every year and in all grades. The small classes made the biggest difference in the scores of children in inner-city schools. Researchers are still following students who participated in the experiment, and they've found that the benefits of small classes in the early years last at least into high school--long after students are back in regular-size classes.

Achilles, C.M., Finn. J.D., & Bain. H.P. (1997/98). Using class size to reduce the equity gap. *Educational Leadership, 55* 940, 40-43.

Perry Preschool Program

Four-year-old children were assigned at random to attend an enriched preschool program or stay at home. The preschool curriculum emphasized active learning, in which the children engage in activities that involve making choices and problem solving, and plan their own activities, carry them out, and review them, with support and encouragement from adults. The study found that compared to the control groups, by age 27, the program increased the proportion of individuals that had graduated from high school or received General Education Development certification, quadrupled the proportion earning more than $2000 per month, reduced the proportion which had ever been on welfare by one quarter, and reduced the percentage of hard-core criminals by four fifths.

Berrueta-Clement, J.R., Schweinhart, L.J., Barnett, W.S., Epstein, A.S., & Weikart, D.P. (1984). *Changed lives.* Ypsilanti, MI: High/Scope.

James Comer School Development Project (SDP)

Incorporated into the program are three mechanisms (the governance and management team, student support team, and the parent program), three guiding principles (collaboration, consensus, and no fault problem-solving), and three operations (comprehensive school plan, staff development, and periodic assessment and modification). Students were randomly assigned to a school using SDP or to a school keeping the current program. Based on data on randomly selected students in matched schools, students in SDP schools achieved significantly higher averages in mathematics and overall grades and in reading, mathematics, and language scores on the California Achievement Test. In some cases, standardized test scores were not significantly higher.

Cook, T.D., Habib, F., Phillips, M., Settersten, R.A., Shagle, S. & Degirmencioglu, M. (1999). Comer's school development program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal, 36* (3), 543-597.

National Reading Panel Studies

In 1997, Congress oversaw the convening of a national panel to assess the effectiveness of different approaches used to teach children to read. For over two years, this National

Reading Panel (Panel) reviewed research-based knowledge on reading instruction. The rigorous research methodological standards used to screen and select the studies that were reviewed and included in the report conform to the "high-quality" standards set forth by IES and are viewed as a model for other large scale curriculum and methological studies. The screening process identified a final set of experiment or quasi-experiment research studies that were then subjected to detailed analysis. The Panel concludes that all students must be taught alphabetics (phonemic awareness and phonics), reading fluency, vocabulary, and strategies for reading comprehension. The Panel endorses systematic phonics instruction for all students and one-to-one tutoring for struggling readers. The panel withholds its approval of sustained silent reading and other practices that the reading research studies selected do not show to be significantly effective and generalizable for all students.

National Reading Panel. (2000). Teaching *children read*. Washington, DC. National Institute of Health and Human Services. Available online at http://www.nichd.nih.gov/publications/nrp/findings.htm

Appendix C

Annotated Bibliography of Leading Educational Technology Research Groups and Representative Research from these Groups

Association for the Advancement of Computing in Education (AACE)

Norfolk, VA. http://www.aace.org

The Association (founded in 1981) is an international, educational and professional not-for-profit organization dedicated to the advancement of the knowledge, theory, and quality of learning and teaching at all levels with information technology. The AACE Digital Library is a valuable online resource on the latest research, developments, and applications related to all aspects of Educational Technology and E-Learning. AACE publishes a number of useful journals relating to computing in education.

Sherry, L., Jesse, D., & Billig, S. (2002). Creating a WEB of Evidence of Student Performance in A Technology-Rich Learning Environment. *International Journal on E-Learning 1*(1), 33-42. [Online]. Available: http://dl.aace.org/6528

Center for Applied Special Technology (CAST)

http://www.cast.org/

CAST is a not-for-profit organization that uses technology to expand opportunities for a people, especially those with disabilities. The report reference below contains the history, research methods, findings and conclusions from a multi-year study of the use of technology to support reading activities between caregivers and children (pre-K and K-3) in family literacy centers. The project resulted in a replicable model that is described in the report.

Coyne, P. & Hughes, B. (1999). *Family and Community Literacy Final Report.* Available online at http://www.cast.org/udl/index.cfm?i=335

Center for Applies Research in Educational Technology (CARET)

Eugene, OR. http://caret.iste.org

CARET is a project of ISTE in partnership with Education Support Systems. CARET was founded in 2000 with a grant from the Bill & Melina Gates Foundation. CARET provides online article reviews, a reading list, helpful resources, and a glossary. The list of online journals is particularly helpful as well as research evidence on how technology can influence student academic performance organized by six research-based dimensions of technology and learning.

Boster, F. J., Meyer, G. S., Roberto, A. J., & Inge, C. C. (2002). *A report on the effect of the unitedstreaming(TM) application on educational performance*. Farmville, VA: Longwood University. http://caret.iste.org/index.cfm?fuseaction=studySummary&StudyID=852&words=Boster&from=searchStudiesKeyword

Center for Children and Technology (CCT)

New York, NY http://www.edc.org/CCT/ccthome/

CCT is a division of the Education Development Center an international, non-profit organization with 325 projects dedicated to enhancing learning, promoting health, and fostering a deeper understanding of the world. CCT investigates how technology can make a difference in children's classrooms, school and communities. Originally founded at Bank Street College and located in New York, CCT conducts basic, applied and formative research as well as designs and develops multimedia tools and technology prototypes.

Carrigg, F. & Honey, M. (2003). *No miracles here—a multi-year study of the use of technology to support reading in family literacy centers*. Available online at http://www2.edc.org/cct/publications_feature_summary.asp?numPubId=127

Center for Highly Interactive Computing In Education (hi-ce) a

Ann Arbor, MI http://www.hi-ce.org

Hi-ce is a group of educators, computer scientists, psychologists, scientists, and learning specialists at the University of Michigan dedicated to educational reform through inquiry-based curricula, learner-centered technologies, comprehensive professional development, and administrative and organizational models. Hi-ce helps schools and teachers change through the infusion of innovative curriculum materials and cutting-edge technologies. Intensive research efforts help to constantly refine and adapt the Center's products to meet the diverse needs of schools, teachers, and students.

Stratford, S. (1996). *A review of computer-based model research in precollege science classrooms.* Ann Arbor, MI: Center for Highly Interactive Computing in Education. Available online at http://www.hi-ce.org/hiceinformation/index.html

Center for Technology in Learning (CTL), SRI International.

Menlo Park, CA http://ctl.sri.com/index.jsp

CTL's mission is to improve learning and teaching through innovation and inquiry in computing and communications. CTL research and development seeks to advance theory and research on effective learning and teaching, and to embody these insights in the innovative design, use, and assessment of interactive learning environments. Much of the Center's work is conducted in real educational settings, such as classrooms and teacher education programs. CTL is a division of SRI International, an independent, nonprofit research institute that specializes in creating and evaluating technology solutions.

Huntley, M.A., Zucker, A., Estey, E. (2000). *A review of research on computer-based tools (spreadsheets, graphing, data analysis, and probability tools), with an annotated bibliography.* Arlington, VA: SRI International. Available online at http://www.sri.com/publication/

Consortium for School Networking (CoSN)

Washington, DC http://www.cosn.org/

CoSN is a national non-profit organization with a mission to advance the K-12 education community's capacity to effectively use technology to improve learning through advocacy, policy and leadership development. Members represent school districts, state and local education agencies, nonprofits, companies and individuals who share the CoSN vision.

CoSN. (2004) *New Studies -- Pro & Con -- on Computers/Internet and Schools*.Available online at http://www.cosn.org/resources/091400a.htm

enGauge by NCREL and the Metri Groups identifies 21st century skills, analyzes the conditions essential for effective technology, hosts an online assessment tool for schools and districts, profiles research-based learning strategies that work.

enGuage ,NCREL, & Metiri Group. (2002). *21st Century Skills—what students need to thrive in today's digital age.* Available online at http://www.ncrel.org/engauge/skills/skills.htm

Institute for Transfer of Technology to Education (ITTE)

Alexandria, VA

http://www.nsba.org/site/page_micro.asp?TRACKID=&CID=63&DID=195

Education Technology Programs was launched in 1985 by the National School Boards Association (NSBA) and its federation of state school boards associations to help advance the wise use of technology in public education. ITTE advances the shared vision by empowering education, industry, and policy leaders to improve education processes and student achievement through knowledge and understanding of technology and organizational development. Its membership component the **Technology Leadership Network** (TLN), founded in 1987, is a consortium of over 400 school districts. Through TLN membership, school districts share ideas and strategies for using technology to transform teaching and learning in the classroom, improve administrative practices to implement significant organizational change, and find new ways to connect with the community. The *Toolkit* is a collection of resources -- research reports, case studies, discussion guides and other tools -- the concepts of which continue to ring true in technology integration.

NSBA. (1997). *Education leadership tool kit—change and technology in America's schools.* Alexandria, VA: author. Available online at http://www.nsba.org/sbot/toolkit/

International Society for Technology Education (ISTE)

Washington, DC. http://www.iste.org/

ISTE is a nonprofit professional society. Besides its work on CARET, ISTE hosts an annual National Educational Computing Conference, and publishes educational technology books and journals, including *Learning and Leading with Technology*, *Journal of Research on Technology in Education*, and *Journal of Computing in Teacher Education*. ISTE's National Education Technology Standards (NETS) defines indictors of effective technology integration for students, teachers, administrators, and teacher education programs.

ISTE (2003). *Technology Foundation Standards for All Students, Educational Technology Standards and Performance Indicators for All Teachers, Educational Technology Standards and Performance Indicators for Administrators.* Washington, DC: author. Available online at http://cnets.iste.org/students/s_stands.html; http://cnets.iste.org/teachers/t_stands.html; http://cnets.iste.org/administrators/a_stands.html

International Technology Education Association (ITEA)

Blacksburg, VA http://www.iteawww.org/

ITEA is the largest professional educational association, principal voice, and information clearinghouse devoted to enhancing technology education through experiences in schools (K-12). Its membership encompasses individuals and institutions throughout the world with the primary membership in North America. ITEA conducts various professional development programs and holds an annual conference. ITEA publishes *The Technology Teacher, Technology and Children, The Journal of Technology Education*, and a variety of other publications and videos that lead the profession by providing teaching directions, instructional ideas, and networking opportunities.

ITEA. (2004). *Measuring Progress: A Guide to Assessing Students for Technological Literacy.* Reston, VA: author. Available online at http://www.iteawww.org/

The **Metiri Group**

Culver City, CA. http://www.metiri.com

The Metiri Group is an educational consulting group with services in policy consulting, professional development, research and evaluation. The firms' services are designed to empower educators and education institutions to: advance effective teaching and learning, use technology in powerful and meaningful ways, and foster 21st Century Skills in students, teachers and administrators. Metiri has developed a set of tools in direct response to client needs, i.e., surveys aligned to national or state frameworks or standards that are customized for a particular evaluation or consulting task.

Metiri Group (2003). Metiri Tools. Culver City, CA: Author. Samples available online at http://tools.metiri.com/survey/samples/

The **Miliken Exchange on Education Technology**

Santa Monica, CA http://www.mff.org

The Miliken Family Foundation has sponsored a number of technology reports and initiatives. The Miliken Exchange conducted an extensive research review of students using technologies including: computer assisted instruction, integrated learning systems, simulations and software that teaches higher order thinking, collaborative networked technologies, and design and programming technologies. These studies show that in over 700 empirical research studies students with access to one of the afore mentioned technologies show positive gains in achievement on researcher constructed tests, standardized test, and national tests.

Schacter, J. (1999). *The impact of education technology on student achievement-What the most current research has to say.* Saint Monica: Miliken Exchange on Education Technology. Available online at http://www.mff.org/pubs/ME161.pdf

North Central Regional Technology Education Consortium (NCRTEC)

Naperville, Il. http://www.ncrel.org

NCRTEC is one of 10 Regional Technology in Education Consortia (RTEC) founded by the Office of Elementary and Secondary Education of the U.S. Department of Education. Among the RTECs, NCRTEC houses particularly good references on research in educational technology. The NCRTEC mission is to help schools and adult literacy programs to develop technology-embedded practices that lead to improved and engaged learning for students. NCRTEC provides resources on professional development, technology planning and evaluation, and tools to enhance effective teaching and learning. Their service region spans seven states: Illinois, Indiana, Michigan, Iowa, Minnesota, Ohio, and Wisconsin.

NCRTEC. (2004). Critical issue: Using technology to improve student achievement. *Pathways.* Naperville, Il: author. Available online at http://www.ncrel.org/sdrs/areas/issues/methods/technlgy/te800.htm

The Sloan Consortium (Sloan-C)

Needham, MA http://www.sloan-c.org

Sloan-C encourages the collaborative sharing of knowledge and effective practices to improve online education in learning effectiveness, access, affordability, and user satisfaction. Sloan-C has found that outcomes from online learning are at least as good as face-to-face learning outcomes.

Bourne, J. & Moore, J. (2003). *Elements of quality online education.* Needham, MA: Sloan Consortium. Available online at http://www.sloan-c.org

Allen, I.E. & Seaman, J. (2002) *Sizing the opportunity—The quality and extent of Online Education in the United States, 2002 and 2003.* Needham, MA: Sloan Consortium. Available online at http://www.sloan-c.org

Pew Internet Project

San Francisco, CA http://www.pewinternet.org

The Pew Internet & American Life Project is a non-profit initiative of the Pew Research Center for People and the Press. Support for the project is provided by The Pew Charitable Trusts. The basic work-product of the center is phone and online surveys; data-gathering efforts that will often involve classic shoe-leather reporting from government agencies, academics, and other experts; fly-on-the-wall observations of what people do when they are online; and other efforts that try to examine individual and group behavior. The Project intends to release 15-20 pieces of research a year, varying in size, scope, and ambition. The Pew Internet & American Life Project creates and funds original, academic-quality research that explores the impact of the Internet on children, families, communities, the work place, schools, health care and civic/political life. The Project aims to be an authoritative source for timely information on the Internet's growth and societal impact, through research that is scrupulously impartial.

Pew Internet and American Life Project. (2002). *The digital disconnect: The widening gap between Internet-savy students and their schools.* Washington, DC: Pew Research Center. Available online at http://www.pewinternet.org/reports/toc.asp?Report=67