This document contains both chapters within this book, co-authored by Dr. Linda M. Phillips:

Chapter 7

Considering Research Quality and Applicability Through the Eyes of Stakeholders

Denyse V. Hayward and Linda M. Phillips pp. 139-148

and

Chapter 27

The Gold Standard and Knowing What to Do

Stephen P. Norris, Linda M. Phillips, and John S. Macnab pp. 603-620

Chapter 7
Considering Research Quality and Applicability Through the Eyes of Stakeholders

Denyse V. Hayward and Linda M. Phillips

Quality in educational research and practice has come under considerable scrutiny by policy makers in the United States. This scrutiny is due, in part, to a desire to develop and implement efficient and effective interventions based on scientific evidence and, in part, by concerns that investment in practices that lack adequate empirical support may drain limited resources. Consequently, there has been a move toward the adoption of the evidence-based practice (EBP) model and accompanying evidence hierarchies from medicine by policy makers and funding agencies as a means to evaluate the quality of education research and to allocate research funding. It is imperative for any discussion of the EBP model in education to know the model as it was conceptualized and implemented in medicine. Sackett, Rosenberg, Gray, Haynes, and Richardson (1996) described EBP in medicine as "the conscientious, explicit and judicious use of best current evidence in making decisions about the care of individual patients" (p. 71). Implementation of an EBP model in medicine involves five essential steps:

1. Convert information needs into answerable questions (formulate the problem).
2. Track down, with maximum efficiency, the best evidence with which to answer these questions—evidence may come from clinical examination, the diagnostic laboratory, published literature, or other sources.
3. Appraise the evidence critically (weigh up) to assess its validity (closeness to the truth) and usefulness (clinical applicability).
4. Implement the results of the appraisal in clinical practice.
5. Evaluate performance. (Greenhalgh, 2006, p. 2)

Adopting and implementing EBP requires that practitioners not only read research but also read the research at the right time and alter their clinical behaviors and the behavior of others in light of what they have found (Greenhalgh). Hierarchies have been developed to support practitioners' critical appraisal and trustworthiness of the research evidence. In evidence hierarchies that evaluate quantitative research designs, studies that conduct systematic reviews of randomized controlled trials (RCTs) and studies utilizing RCTs are at the pinnacle (Greenhalgh). Thus, the EBP model is appealing because it appears to offer objective criteria to determine

D.V. Hayward and L.M. Phillips
University of Alberta

M.C. Shelley II et al. (eds.), *Quality Research in Literacy and Science Education*,

best practice (Horner et al., 2005) since it allows for types and strengths of *evidence* to be differentiated.

There is considerable debate regarding the appropriateness and applicability of adopting EBP and the accompanying evidence hierarchies from medicine to education. Participants at the 2nd Island Conference discussed many of these issues, and the authors of Part I of this book discuss the implementation of EBP from a variety of perspectives. Our goal in this chapter is to highlight and discuss important concepts and issues raised by these authors as they relate to various stakeholders.

7.1 Evidence-based Practice—What Counts as Evidence?

Yore and Boscolo (see Chap. 2) began by situating the issues that are discussed in each chapter within the broader context of the shift toward EBP and legislation—Gold Standards in Education Research (Bush, 2002); No Child Left Behind Act of 2001 (NCLB, 2002)—for education research. This shift is described by the authors as a result of (a) ideological and political agendas to improve educational outcomes for all students and (b) skepticism regarding the quality, rigor, and effects of research effectiveness on student outcomes. Yore and Boscolo discuss the challenges that have resulted from misunderstandings or misinterpretations in the translation of legislation that has privileged quantitative methodologies and evidence hierarchies, in particular RCTs, rather than focusing on research designs (quantitative, qualitative, mixed methods) that are appropriate to answer particular research questions. Stakeholders at different levels of the implementation process will have differing but important perspectives regarding EBP that other stakeholders need to consider, address, and incorporate.

7.1.1 Educators, Employers, and Professional Bodies

Many of these stakeholders rightfully question whether EBP, like so many other practices of the past, is just the latest fad. Upon hearing that EBP challenges them to consider questions such as *How do you know that what you do works?* many teachers indicate that they regularly ask such questions because it is part of what constitutes good teaching practice. However, proponents of EBP state that what sets EBP apart is the emphasis on using scientific evidence to answer such questions rather than relying on expert opinion or past practice (Greenhalgh, 2006; Reilly, 2004). Proponents argue that by adopting an EBP model educators will be more able to critically appraise the benefits and risks associated with particular instructional methods, interventions in classrooms, and individual student contexts.

Problematic to the claims made by EBP proponents is the lack of consensus for the EBP model across any discipline, including medicine (see Beecham, 2004;

Greenhalgh, 2006; Odom, Brantlinger, Gersten, Horner, Thompson, & Harris, 2005). For many nonmedical disciplines, the conceptualization and underlying assumptions of the evidence-based medicine model are at odds with the conceptualization and reality of their practitioner–patient or teacher–student relationships. Beecham spoke to this issue as it relates to the discipline of speech–language pathology. She argued that speech–language pathologists (SLPs) understand their practice differently from that of medical practitioners. For SLPs, the establishment of equitable and collaborative practitioner–patient relationships is viewed as central to, and an important component of, the success of therapeutic goals. Thus, the EBP model adopted from medicine, where evidence focuses only on external, measurable variables, is problematic. Many of the variables that support success in a collaborative treatment context are neither external nor easily measured. Given that a large proportion of speech–language pathology practice occurs within educational contexts, Beecham's arguments are informative and insightful for educators.

Recommendations made by EBP proponents, however, are often presented as though there is consensus as to what counts as evidence and what sorts of evidence are better than other evidence (Johnston, 2005). Johnston noted that amidst the enthusiasm for EBP it is easy to lose sight of the fact that these assumptions are virtually untested when adopted by other disciplines, often left unstated, and most definitely arguable, as shown by Beecham (2004). With the existence of considerable and substantial debate within and across disciplines, it is reasonable for educators, employers, and professional bodies to be confused about why EBP should be adopted—given that the costs of such change are substantial for this particular set of stakeholders.

7.1.2 Policy Makers and Funding Agencies

In their zeal to be fiscally responsible, policy makers and funding agencies' stakeholders need to carefully weigh the available evidence that exists in the research literature across a number of disciplines that have attempted to adopt EBP from medicine. Legislation of a practice model that will have substantive human and financial costs requires a priori knowledge of known problems in the conceptualization of the particular model. It is clear from a variety of publications (see Graham, 2005), however, that conceptual clarity has not been achieved; unfortunately, practitioners and researchers with the least power to affect change in ill-conceived and poorly articulated policies are left to face the consequences.

7.2 Uptake of Research Evidence

Millar and Osborne (see Chap. 3) begin by citing comments made by Hargreaves (1996) that educational research has offered little to inform teaching practice over the past 50 years because research studies are noncumulative, produce inconclusive

and contestable findings, and are of little practical relevance. This position appears to have some support amongst practitioners (e.g., Lijnse, 2000) who have expressed dissatisfaction with the lack of research evidence to support teaching. Millar and Osborne devote the remainder of their chapter to examining this research-to-practice issue within the context of EBP. Three actual examples of instructional approaches in the teaching of science that are cumulative and conclusive—and have substantive practical relevance—are presented. Although all three studies had significant impact for the schools in which the research was conducted, broader application in science teaching has not occurred for at least two of these approaches. EBP proponents would argue that the lack of broad impact relates to the weakness of the evidence these studies offer because none were conducted using RCT designs. However, Millar and Osborne examined such a claim and concluded that it would be difficult to justify the expense in human and material resources to achieve the same findings using a RCT methodology for the three examples cited.

The reluctance to engage in, indifference toward, or ignorance of research evidence for purposes of uptake is of considerable importance for all stakeholders. Although Millar and Osborne demonstrate that it is clearly not simply a void in the availability and accumulation of quality research evidence, as suggested by Hargreaves (1996), there is limited expectation on the part of practitioners and policy makers that relevant research exists and an even lower expectation that research is to inform policy and practice. Such perceptions persist at all stakeholder levels and must be addressed if we are to make advances.

7.2.1 Educators

Sweeping statements, such as those made by Hargreaves (1996), denigrating the relevance of education research have serious consequences. First, such comments permit educators and others to dismiss relevant research findings out of hand. Second, such comments diminish the significant advances made in literacy and science education research. Finally, once such disregard is permissible, it becomes even more difficult to convince educators that any model, including EBP, will improve circumstances. Many authors throughout this book have reported on, referred to, and mentioned relevant and important research in literacy and science education that has left each of us with a greater appreciation of how our individual research fits within the larger picture of education—a picture that differs little from other areas in the social sciences and humanities.

If, as proponents suggest, the EBP model holds promise in bridging the gap between research evidence and practitioner uptake for the field of education, then the question remains as to how educators are to develop the skills necessary to implement an EBP model in classrooms in order to take advantage of research-based evidence to teach particular content, grade, and developmental levels. Many articles, chapters, and books (e.g., Greenhalgh, 2006; Johnston, 2005; Reilly, 2004; Silagy & Haines, 2001) are devoted to outlining the skills practitioners across a variety of

disciplines need to develop in order to implement EBP. For example, the following skills are offered by Reilly: (1) completing a course or online tutorial on EBP, (2) developing critical appraisal skills when reading research papers, (3) becoming skilled users of research to enable the application of scientific information in their day-to-day practice, and (4) developing questions related to day-to-day practice that can be answered using evidence-based research. Unfortunately, educators often find themselves having to undertake learning skill sets such as those described with minimal or no support from employers, professional bodies, or the government agencies mandating practice changes. Many educators question whether the time needed to learn new skills, often at their own expense, is worth it, if EBP will likely be replaced in an ever-changing political agenda.

7.2.2 *Employers, Professional Bodies, Preservice Education Programs, Funding Agencies, and Policy Makers*

EBP proponents advocate and purport that research conducted using RCT will improve research uptake in education practice; however, evidence from medicine and other health professions does not support this contention. Many examples exist where evidence from RCTs demonstrated that particular interventions are not beneficial and may even be detrimental, yet these interventions continue to be widely used (see Gillam, Crofford, Gale, & Hoffman, 2001, for *Fast ForWord* language intervention; Greenhalgh, 2006, for back pain; Phillips, Norris, & Steffler, 2007, for *Meaningful Applied Phonics* reading instruction). Odom et al. (2005) suggested that EBP proponents have ignored the issue of whether or not results from RCTs are positive.

Further, there is evidence showing that, while health care practitioners consider research to be important, research findings have little impact on their day-to-day practice (Brener, Vallino-Napoli, Reid, & Reilly, 2003; Metcalfe et al., 2001). Reilly (2004) found that practitioners tend to read the abstract, introduction, and discussion sections of research articles but feel much less confident about understanding methods and results sections. Yet, to conduct critical appraisals of the research literature, these are the very sections that educators need to understand. If such is the modus operandi amongst the health profession that have implemented EBP for a much longer period of time, then we must question whether we realistically can expect a different outcome in education.

Logemann (2004) pointed to yet another issue that impacts uptake of research evidence, that is, the focus on productivity in health care and educational institutions. A productivity model is at odds with EBP, which requires time to develop expert skills, acquire new knowledge, and read and apply evidence. Currently, the cost of developing expert skills is not included in funding models in health care (Reilly, 2004) or education, but is an important issue for these stakeholders to consider if the EBP model is to be adopted in education consistently and successfully.

7.2.3 Researchers

Uptake of research evidence by educators is a significant concern for researchers. Researchers can support not only practitioners but also audiences across all levels if, according to Johnston (2005), there is a concerted effort to (a) situate the research within the larger context of the problem being studied, (b) provide clear indications for educational practice, and (c) clearly explain the extent of any limitations or generalizability issues. Logemann (2004) also suggested that researchers take the lead by conducting systematic reviews of assessment and intervention strategies as a means to critically appraise and synthesize the research literature for specific issues. Such syntheses, according to Logemann, would be helpful to practitioners who have limited time and resources to access and examine the available research. However, this recommendation would mean examining studies across a much broader range of methodologies than is the current practice (Johnston). We would add that, unless issues of why practitioners do not use research in practice contexts are addressed by all stakeholders, no improvement in uptake of research information is likely to occur no matter how exhaustive or clearly written the information.

7.3 Misinterpretation of Evidence Hierarchies

Two chapters in Part II focus on demonstrating the limitations of the wholesale adoption of evidence hierarchies developed for medicine to determine strengths of evidence in educational research and the allocation of research funds. Alvermann and Mallozzi (see Chap. 4) highlight the contributions of qualitative and quantitative research perspectives to teaching and learning, while Tytler (see Chap. 5) presents evidence from longitudinal studies showing that RCTs can neither duplicate nor supplant important insights yielded by these designs. The important issue raised by these authors relates to policy implementation, where misinterpretations of particular research methodologies are sanctioned whilst others are discouraged and denied funding for research programs. The consequence of misinterpretation narrows not only the range of questions that can be researched but the type of information that will be available to educators to support teaching and learning.

7.3.1 Policy Makers and Funding Agencies

The appeal of RCT design is that it reduces bias and increases generalizability of results because treatment groups are equivalent and representative of the larger group with the exception of the intervention received. Even in medicine, where RCTs are considered the Gold Standard, problems exist in optimal implementation. Due to the expensive, time-consuming nature of RCTs, many studies are conducted with inadequate numbers of participants or too short a time frame (Greenhalgh, 2006).

She added that there are often hidden biases in RCTs that result from imperfect randomization, failure to randomize all applicable individuals, and failure to blind examiners to the randomization status of study participants. Exclusion and inclusion biases also limit generalizability of RCT findings. In education, individuals with learning or reading disabilities, low socioeconomic status, behavioral or attention difficulties, or from minority populations are often excluded. The *normal* participants in many RCT study samples will likely differ in important ways from students within a particular school or community thus confounding results and limiting generalizability (Montgomery & Turkstra, 2003). The heterogeneity of participant characteristics and individuals with low-prevalence disorders and disabilities—as is common in educational contexts—poses a significant challenge to RCT research designs, which are based on establishing equivalent groups and where relatively large numbers of participants are needed to achieve analytical power (Greenhalgh).

These are all important considerations that have been overlooked in the shift of emphasis to RCT designs to the exclusion of other designs in education. However, by far the most significant problem overlooked by the RCT shift in funding allocation is that RCT designs are *only* applicable to questions regarding intervention. RCTs are not appropriate to answer questions related to diagnosis, prognosis, motivation, preferences, or beliefs; examination of these important issues requires quantitative, qualitative, and mixed-method designs (Greenhalgh, 2006). Excluding or limiting the pursuit of these critically important issues goes directly against the purpose of the legislation.

7.4 High-quality Research Requires Adequate Funding Support

The penultimate chapter in Part II (see Chap. 6) offers a review of mechanisms used to evaluate quality in education research across seven nations: Australia (AU), Brazil (BR), New Zealand (NZ), Singapore (SG), South Africa (ZA), Taiwan (TW), and the United Kingdom (UK). The authors found that mechanisms were dependent on the overarching aim of education for each nation; these included: (a) accountability of public funds (AU, NZ, UK), (b) improvement in economic performance and quality of life (NZ, SG, ZA, TW), and (c) making educational institutions comparable to institutions internationally (BR). Aims across nations were similar to those in the USA; however, no particular research methodology was privileged by any of the seven nations.

All countries identified constraints in developing and conducting high-quality research programs. The range of constraints included: (a) lack of government-level financial support resulting in numerous high-quality projects failing to be funded, administrative burden, and legislative demands (AU, BR, NZ, ZA, TW, UK); (b) lack of expertise and human resources to conduct research (SG); (c) cultural and racial issues related to the apartheid regime (ZA); and (d) reluctance by schools to be involved in educational research (BR).

The international survey revealed a clear commitment to quality in educational research but a consistent lack of funding to support high-quality research programs. These issues require the attention of policy makers and funding agencies.

7.4.1 Policy Makers and Funding Agencies

Chapter 6 by Coll and colleagues speaks to an international commitment to the application of quality indicators that represents rigorous application of research methodologies appropriate to answer the particular questions. Such indicators serve as guidelines for (a) researchers designing and conducting research, (b) policy makers and funding agencies evaluating the believability of research findings, and (c) educators determining the usability of research findings (Horner et al., 2005).

All seven nations achieve high-quality research without an emphasis on particular methodologies. In fact, Coll and colleagues show that relatively few countries are even in a position to conduct large-scale projects that might lend themselves to RCT research designs. Additionally, the expense of such studies would be problematic for the majority of nations. It is clear across all nations that the lack of financial resources available from government and funding agencies impacts both development and implementation of high-quality research programs. If policy makers and funding agencies are serious about committing to improving educational outcomes for all students, then increased financial support for research programs, including RCTs, is needed.

7.5 Conclusions and Implications

Berliner (2002) proposed that scientific research in education is not a hard science—such as medicine, chemistry, and biology—but it is the hardest-to-do science. Educational researchers conduct scientific research under conditions that physical scientists would find intolerable. They face particular problems and must deal with local conditions that limit generalizations and theory building—problems that are different from those faced by the easier-to-do sciences of chemistry, biology, and medicine. Mandating EBP has a significant impact on stakeholders at all levels. When there is less than optimum understanding and acceptance of new practice models, consistent and successful implementation is seriously challenged.

One of the two prominent issues raised by the authors in Part II is the appropriateness of the wholesale adoption of an EBP model and accompanying evidence hierarchies developed for medical practice to educational practice. The assumptions of the EBP model are virtually untested when adopted by other disciplines, frequently left unstated, and most definitely arguable (Johnston, 2005). The potential danger of focusing more or less solely on EBP is that it leads to disproportionate emphasis

on the tools of experimental design rather than the specific questions that need to be answered (Montgomery & Turkstra, 2003). Greenhalgh (2006) concurred, stating:

> [W]hen applied in a vacuum (that is, in the absence of common sense and without regard to the individual circumstances and priorities of the person being offered treatment) the evidence-based approach to patient care is a reductionist process with a real potential for harm. (p. xiii)

A unidimensional focus on funding RCTs in intervention research in education is misinformed. By adopting such a position, the implication is that only intervention studies are needed to support teaching and learning. Studies engaged in diagnosis, screening, prognosis, and motivation—all of which most stakeholders consider imperative to the success of both teaching and learning—could not be conducted since RCT is an inappropriate methodological choice. We propose that if policy makers and funding agencies had enacted the five essential steps to implement evidence-based practice then many of the problems in adopting the model in education may have been preempted.

The other prominent issue concerns lack of uptake of research evidence in educational practice. This is a complex issue with a variety of reasons posited, including: (a) practitioners claim that there is a lack of any research to support practice, (b) research participants or treatments do not represent the reality in everyday practice, and (c) lack of time to access research evidence. The acknowledgment that educational practice functions primarily as a productivity model, which is at odds with the EBP model, is a significant consideration for all stakeholders since the development of these EBP skills is not included in funding models. We suggest that government policy is also more closely aligned to a productivity model, which is also at odds with the mandated legislation.

Despite the initial difficulties, we strongly believe that stakeholders in education have the opportunity to be leaders in developing an evidence model and accompanying hierarchies. Such developments within education that adequately address the types of research that best take account of the complexities of conducting educational research and the numerous challenges faced by educators in the uptake of research evidence are necessary for and fundamental to the education of our nations' children.

References

Beecham, R. (2004). Power and practice: A critique of evidence-based practice for the profession of speech-language pathology. *Advances in Speech Language Pathology, 6*(2), 131–133.

Berliner, D. C. (2002). Educational research: The hardest science of all [Comment]. *Educational Researcher, 31*(8), 18–20.

Brener, L., Vallino-Napoli, L. D., Reid, J. A., & Reilly, S. (2003). Accessing the evidence to treat the dysphagic patient: Can we get it? Is there time? *Asia Pacific Journal of Speech, Language and Hearing, 8*(1), 36–43.

Bush, G. W. (2002, November 5). *Statement on signing legislation to provide for improvement of federal education research, statistics, evaluation, information, and dissemination, and for*

other purposes. Retrieved February 14, 2008, from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname = 2002_presidential_documents&docid = pd11no02_txt-21.pdf

Gillam, R. B., Crofford, J. A., Gale, M. A., & Hoffman, L. M. (2001). Language change following computer-assisted language instruction with *Fast ForWord* or *Laureate Learning Systems* software. *American Journal of Speech-Language Pathology, 10*(3), 231–247.

Graham, S. (Ed.). (2005). [Special Issue]. *Exceptional Children, 71*(2).

Greenhalgh, T. (2006). *How to read a paper: The basics of evidence-based medicine* (3rd edn.). London: Blackwell BMJ Books.

Hargreaves, D. H. (1996). *Teaching as a research based profession: Possibilities and prospects* [The Teacher Training Agency Annual Lecture]. London: The Teacher Training Agency.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179.

Johnston, J. R. (2005). Re: Law, Garrett, & Nye (2004a). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language & Hearing Research, 48*(5), 1114–1117.

Lijnse, P. L. (2000). Didactics of science: The forgotten dimension in science education research. In R. Millar, J. Leach, & J. Osborne (Eds.), *Improving science education: The contribution of research* (pp. 308–326). Buckingham, UK: Open University Press.

Logemann, J. (2004). Evidence-based practice. *Advances in Speech Language Pathology, 6*(2), 134–135.

Metcalfe, C., Lewin, R., Wisher, S., Perry, S., Bannigan, K., & Moffett, J. K. (2001). Barriers to implementing the evidence base in four NHS therapies: Dietitians, occupational therapists, physiotherapists, speech and language therapists. *Physiotherapy, 87*(8), 433–441.

Montgomery, E. B., Jr., & Turkstra, L. S. (2003). Evidence-based practice: Let's be reasonable. *Journal of Medical Speech Language Pathology, 11*(2), ix–xii.

No Child Left Behind Act of 2001. Pub. L. No. 107–110, 115 Stat. 1425. (2002).

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*(2), 137–148.

Phillips, L. M., Norris, S. P., & Steffler, D. J. (2007). Potential risks to reading posed by high-dose phonics. *Journal of Applied Research on Learning, 1*(1), Article 2, 1–18.

Reilly, S. (2004). The challenges in making speech pathology practice evidence based. *Advances in Speech Language Pathology, 6*(2), 113–124.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*(7023), 71–72.

Silagy, C., & Haines, A. (Eds.). (2001). *Evidence-based practice in primary care*. London: Blackwell BMJ Books.

Chapter 27
The Gold Standard and Knowing What to Do

Stephen P. Norris, Linda M. Phillips, and John S. Macnab

The call for evidence-based educational practice presumes that science is a way to good knowing and often presumes as well that good knowing leads more or less directly to good acting. We will not critique science as a means to good knowing, particularly regarding the effectiveness of educational interventions. Rather, we shall urge educators to pay more attention to the relationship between scientific knowledge and what can be done with that knowledge. Providing an accurate view of this relationship is critically important to how science can serve as a vehicle for change in social practice. "At issue are the potency and value ascribed to certain forms of evidence in supporting propositions that arise in educational practice" (Thomas, 2004, p. 1).

Much of the impetus for the recently revived debate about the role of scientific evidence in education stems from two pieces of legislation passed in the United States. The first is the No Child Left Behind Act of 2001 (NCLB, 2002). However, the second, the Education Sciences Reform Act of 2002 (ESRA, 2002), is more important to our business here. ESRA established four new centers in the US Department of Education (US ED): The Institute of Education Sciences, National Center for Education Research, National Center for Education Statistics, and National Center for Education Evaluation and Regional Assistance. Of these, the President of the United States said at a press conference:

> Today I have signed into law H.R. 3801, an act to provide for improvement of Federal education research, statistics, evaluation, information, and dissemination, and for other purposes. This Act will substantially strengthen the scientific basis for the Department of Education's continuing efforts to help families, schools, and State and local governments with the education of America's children. This Act is an important complement to the No Child Left Behind Act enacted earlier this year. (Bush, 2002, para. 1)

It is statements contained in subsequent documents from the Institute of Education Sciences (IES) regarding the use of scientific research in education, to be described presently, that will help to motivate the argument in this chapter,

S.P. Norris, L.M. Phillips, and J.S. Macnab
University of Alberta

M.C. Shelley II et al. (eds.) *Quality Research in Literacy and Science Education*,

which is that no results of research—no matter how well the research is conceived in science or in other forms of inquiry—can by themselves determine what we ought to do in practice.

We have structured the chapter to enrich and extend other chapters in Part V around five sections. In the first section, we provide some additional context and motivation for the problem we wish to address. Section two is devoted to the question of what constitutes good human action—how we know what to do. Section three examines the nature of scientific theories and research-based knowledge in general and explores what *must* be the case for such knowledge to be put into practice. In the fourth section, we look at a particular case of a scientifically based intervention study and demonstrate how problems of implementation arise that have not been contemplated in the documents provided by the IES. Finally, we turn in the fifth section to some conclusions and policy implications.

27.1 Context and Motivation

Science and scientific knowledge frequently are employed to lever change in social practices, such as medical and nursing care, child care and social welfare, and education. In the medical field, the Cochrane Collaboration has for over a decade been dedicated to an increase in scientific evidence-based practice in medicine:

> The Cochrane Collaboration is an international not-for-profit and independent organization, dedicated to making up-to-date, accurate information about the effects of healthcare readily available worldwide. It produces and disseminates systematic reviews of health care interventions and promotes the search for evidence in the form of clinical trials and other studies of interventions. (Cochrane Collaboration, n.d., para. 1)

A quick scan of the Cochrane website (http://www.cochrane.org) reveals that since 1993 the collaboration has produced over 5,000 meta-analyses of medical intervention studies. There has been an annual international colloquium since its inception. Taken together, the meta-analyses and the colloquia have had a staggering effect on medical practice globally.

Although not as organized as the Cochrane Collaboration, there is a similar move toward science in the nursing field. In many jurisdictions, nursing education has moved away from a hospital-based apprenticeship toward a postsecondary institution-based and scientific knowledge-based profession. Obtaining a nursing license typically requires as a minimum a bachelor's degree, and nurses are taught more and more by individuals who have research-based doctorate degrees in nursing. Currently in our own jurisdiction, for example, there is a grave shortage of nurses. However, this shortage is preceded and exacerbated by another, namely, the shortage of Ph.D. nursing professors to teach nursing in the university context. The clear aim in the medical fields has been to base practice squarely upon science, on the presumption that science is a way of good knowing and that such knowledge leads to good practice. In these fields, there is a widespread call to turn to scientific evidence to find out what to do.

Education's history is somewhat different. The move in education toward scientifically based research started early in the 20th century and reached its peak around the middle of the second half of the century. At about the same time as scientific educational research was reaching its most dominant status, trenchant criticisms of science as an objective basis for social science research—criticisms that had been articulated much earlier in the century (e.g., Rudner, 1953)—began to take hold. The effect of these criticisms was a move toward qualitative forms of inquiry and away from experimental control and statistical probability as criteria for educational research conclusions. Some believe that the pendulum has begun to swing back toward scientifically based educational research although most researchers, including ourselves, do not wish to embrace the naive forms of empiricism that typified some research during the 20th century. A clear indication of a swing change was the formation of the Campbell Collaboration in 1999/2000. Whereas the Cochrane Collaboration is named after a famous British epidemiologist (Archie Cochrane), the Campbell Collaboration is named after Donald T. Campbell, who had enormous influence on the conduct of experimental inquiry in education and other social sciences (see, e.g., Campbell & Stanley, 1963). The purpose of the Campbell Collaboration is "to help people make well-informed decisions about the effects of interventions in the social, behavioral, and educational arenas" (Campbell Collaboration, n.d., para. 1). It provides research reviews in the areas of crime and justice, education, and social welfare. In contrast to the thousands of reviews in the Cochrane Collaboration, at the date of writing this chapter, only 5 education reviews were completed by the Campbell Collaboration with another 15 in progress.

A further indicator of the recent shift back toward scientifically based educational research is the legislation in the United States mentioned in the introduction. Shortly after its formation, IES published a document, *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide* (US ED, 2003) that helps frame the argument we make in this chapter (see Shelley, Chap. 22). There are several key features of the document that we will highlight at this point so that they can serve as subsequent foci for critique. The first feature is contained in the title, which indicates that the document is not only about identifying educational practices that are supported by educational research but also about implementing those practices. That is, the document is not only about what we know but also about what to do. We note that, of the total 19 pages in the document, fewer than 1.5 are devoted to implementation. The second feature is the decision to narrow the focus of attention on scientifically based research to "randomized controlled trials—research's 'gold standard' for establishing what works" (p. iii). A third feature is the failure to acknowledge and speak explicitly to how educational goals are adopted and justified. For example, embedded in a statement such as *randomized controlled trials are considered the "gold standard" for evaluating an intervention's effectiveness, in fields such as medicine, welfare and employment policy, and psychology* is a claim or presumption about what counts as effective. No empirical inquiry, no matter how golden, can by itself justify a claim about what counts as effective. Claims about effectiveness and goals are normative and require

for their justification normative arguments, although they may depend also upon relevant empirical evidence.

A fourth, and the final, feature we will highlight is the failure to draw the relationship between knowledge that has been generalized and abstracted from several contexts and the use of that knowledge in particularized and concrete settings. Indeed, in attempting to construct such a relationship, the document falls into a contradiction. First, there is the recognition that "slight differences [between implementation settings and settings of the studies] could lead to substantially different outcomes" (US ED, 2003, p. 14). The suggested way to determine whether different outcomes are occurring in, for example, an evidence-based reading program, is to compare the results in the particular context to "a comparison group of schools or classrooms, roughly matched in reading skills and demographic characteristics, that is not using the program" (p. 14). However, such information, in failing to come from a randomized controlled trial (RCT), would not reach the Institute's lowest level of acceptable evidence; hence, by their own arguments, it cannot override the evidence from the RCTs upon which the reading program, by hypothesis, is based.

The reason for falling into contradiction is that the relationship between Gold Standard evidence and use of that evidence in particular contexts has been drawn too tightly by the IES document. The failure is to recognize that obtaining the evidence and putting it into use are two different activities, each with its distinctive processes of reasoning and justification. Since the activities are different, it is not at all inconsistent to point to a practice as having the best evidence for effectiveness and to decide that the practice is not the right one to adopt in the circumstance, or to adopt a practice that has poor evidence to support it. We will show how the relationship between Gold Standard evidence and its use can be drawn less tightly and, thus, without contradiction, leave room for maneuver in deciding what to do.

As a consequence of the critiques we shall make, we will conclude that education needs to be clear how the results of scientific educational research are related to educational practice. We shall conclude also that scholars of education need to help educational policy makers understand how research can be related to practice.

27.2 What Is Good Acting?

We focus in this section on questions of what we ought to do and what sort of justifications is required to answer satisfactorily such questions. The first point is that good acting and good knowing are intimately connected. What we do somehow must be connected to what we know. If we are concerned with what we *ought* to do, then the knowing had better be *good* knowing. We find support for this seemingly obvious claim from Dewey (1929/1984): "*the* problem of practice is what do we need to *know*, how shall we obtain that knowledge and how shall we apply it" (p. 30). Code (1987) also drew a connection between what we know and how we ought to behave: "an epistemic community will be strong in intellectual virtue

only if good knowing is valued as a condition of human flourishing" (p. 246). Dewey and Code both were speaking of empirical knowledge of which scientific knowledge is the paradigm case. Thus, conclusions based upon the Gold Standard reasonably can be seen as a basis for good acting.

In addition to good empirical knowledge, what we do must also be based upon sound normative principles. We cannot legitimately infer what ought to be done solely from some fact of the matter. As Dewey (1929–30/1984) again said, "laws and facts, even when they are arrived at in genuinely scientific shape, do not yield *rules of practice*" (p. 14). The is/ought gap to which Dewey was alluding is credited to Hume (1739–40/1962) who noticed that one cannot infer without controversy from an *is* to an *ought*, from a description of how things are to a recommendation how they might be. Upon the descriptive statement, *Smoking causes lung cancer*, by itself, one cannot base any of the recommendations:

> The tax on cigarettes should be high enough to discourage smoking.
> Children should not be permitted to purchase cigarettes.
> People should not smoke.

To see how the direct inference fails, simply consider the additional descriptive claim, *Smoking brings pleasure*. It is now apparent that, in order to infer that children or people generally should not smoke or that the tax on cigarettes should be high, it is necessary to show that the negative consequences of smoking outweigh the positive ones. Showing the latter requires normative judgments that no descriptive statements by themselves can settle.

Without introducing some evaluative or normative premise in the form of a sound principle (people should not engage in behavior that causes serious disease; the state should move to discourage people from causing risk to their health; or laws should be enacted that forbid children from engaging in actions that risk grave harm), we cannot infer from what is the case to what we ought to do. Once we suggest such evaluative premises, however, we see that further problems arise in deciding what should be done based upon what we know. Consider the following line of reasoning:

> Smoking causes lung cancer.
> Laws should be enacted that forbid children from engaging in actions that risk grave harm.
> A law should be enacted to prevent children from smoking.

Now, this line of reasoning gets us what we want, that is, the prevention of children smoking. However, the cost of endorsing the particular evaluative premise might be higher than we are prepared to pay. Clearly, we do not wish to enact laws that forbid children from engaging in *all* actions that risk grave harm to them. Else, we would need laws forbidding hockey playing, bicycle riding, roller blading, rope skipping, plugging in the toaster, and perhaps walking down the stairs. The evaluative premise successfully links the descriptive claim that smoking causes lung cancer to the recommendation that children ought not to smoke, but the evaluative claim is too encompassing. Thus, we must seek a narrower claim that still successfully

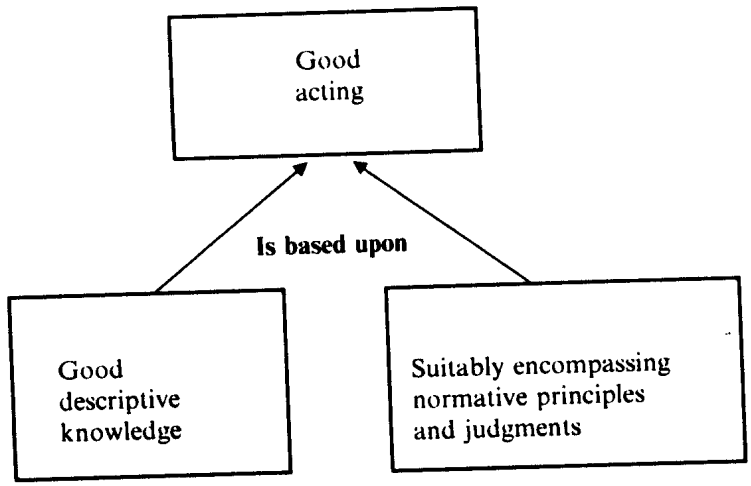

Fig. 27.1 The basis of good acting in good descriptive knowledge and in normative principles and judgments

makes the link between the effects of smoking and our desire to prevent children from doing it but does not so restrict children's lives as to make them unbearable. The evaluative premise might be modified, for instance, to call for laws that forbid risky actions when those actions have no weighty positive outcomes. Such a premise would not rule out bicycle riding; because, even though it carries risks, it also provides enormous benefits in fun, exercise, ease of transportation, and skill and agility development that can last a lifetime.

We thus conclude that good acting requires both good descriptive knowledge—much of which comes from scientific research—and sound evaluative principles, which arise from our imaginations and very broad ethical concepts such as fairness and justice. We have depicted this relationship in Fig. 27.1. The dual basis of good acting is depicted by the dual arrows coming from descriptive knowledge and from normative principles and judgment, both of which serve as part of the basis. To act upon descriptive knowledge alone is a failure in critical reflection, because no amount of knowledge alone can imply what ought to be done. On the other hand, we must be careful of the evaluative premises we choose. If we adopt an evaluative premise to link knowledge to action in one situation, then consistency demands that we apply that premise to all cases that fall under it. If we are not careful in our choice of premises, we can rule out action that we desire highly. Thus, critical reflection is also needed in choosing evaluative premises.

27.3 Nature of Scientific Theories and Research-based Knowledge

In addition to the considerations in the previous section on the link between knowledge and action, the nature of scientific theories also bears upon how they can be put to use. We shall use the term *scientific theory* in our discussion, which

seems acceptable given that the Gold Standard is a norm derived from science. Nevertheless, we do not limit our conclusions to science proper. Indeed, the characteristics of scientific theories that we highlight also are characteristics of research-based knowledge generally, including educational research-based knowledge.

27.3.1 The Semantic Conception of Theories

We derive our view of theories from Suppe (1977, 1989), and a fuller description of our derivation can be found elsewhere (Norris, 2000; Norris & Kvernbekk, 1997). According to Suppe, theories are models or abstract systems. Abstract systems are abstract in the following sense: theories are concerned with phenomena only insofar as the phenomena are characterizable by a small number of parameters abstracted from them. A theory cannot characterize a phenomenon in all of its complexity. Abstraction in this sense must occur if theories are to be general.

Given the attention to science paid by those advocating the Gold Standard, it seems appropriate to examine an example from the natural sciences. Consider the Kinetic Theory of Gases. In addition to the parameters of pressure, volume, and temperature, the theory is based on abstracted parameters for the size of molecules, the shape of molecules, the motion of molecules, the density of molecules, the elasticity of the collisions among molecules and between molecules and the walls of the container, and the attractive and repulsive forces through a distance among molecules and between molecules and the container. Typically, as is the case here, theories contain idealizations on some parameters. We might, for example, idealize an interaction to be negligible or zero, as we do if we assume perfectly elastic collisions among molecules. A state of an abstract system is defined by the values on each of its parameters at a given time, and the behavior of an abstract system is its changes in state over time. Changes in state are defined by laws of the theory.

The nature of the relationship between abstract systems and concrete systems is one of *counterfactuality*. Theories do not describe accurately concrete phenomena but describe what the phenomena *would have* been *had* the selected parameters been the only ones exerting any influence and *had* the idealizations been real. The nature of the relationship between theories and concrete systems leads to a number of implications, which we explore in the following three sections.

27.3.2 Impossibility of Direct Application

The abstraction and idealization necessary for the existence of theories must be taken into account when applying them. There never can be a recipe to get from an abstract and general theory to a concrete and particular system. The connection has to be indirect through auxiliary hypotheses that specify the influence on the concrete system of

factors not specified in the theory. Auxiliary hypotheses are required for *all* applications of abstract systems to concrete systems. Auxiliary hypotheses specify the effects of outside influences that are identified based upon an "appraisal of the situation" (Norris, 2000, p. 181). The possible ways of applying a theory are in principle unlimited, and the possible number of auxiliary hypotheses useful in mediating between a theory and concrete systems is in principle unlimited. For many cases of application, we do not have all of the requisite auxiliary hypotheses, especially not when the concrete system is characterized by great variability and flux. This can be difficult work—deciding how theoretical knowledge needs mediating for use in particular situations.

Theory application always involves normative considerations about whether and how to apply the theory. These normative considerations might involve questions of economics, aesthetics, ethics, and prudence. Such considerations are not part of the theory but may affect making connections between the theory and concrete systems. When applying a theory, there must be some more or less clear notion of what to try to achieve. For example, if the desired accuracy of prediction is low, it perhaps would be sufficient to take into account through auxiliary hypotheses only some of the most important influences that lie outside the scope of the theory. If the desired accuracy is high, then more influences might need to be taken into account, and taken into account more precisely. This final point leads to a discussion of the variability of choice in application.

27.3.3 *Variability in Application*

Situational appraisals cannot be made in the abstract. Rather, they must be made in light of the particular theory being applied, the particularities of the situation, and the outcomes desired. Situational appraisals are *judgments* that can be made only by those knowledgeable of the application situation. Theoretical and practical knowledge have equal importance in the application situation.

Let us consider a situation headlined as follows in a recent newspaper article: "Should your daughter get the needle?" (Anderssen & Alphonso, 2007). The article was about the question of whether or not girls 12–14 years of age ought to be given a new vaccine that has been shown to confer immunity against cervical cancer caused by the human papillomavirus (HPV), a sexually transmitted disease (STD). Here are some of the medical facts:

- There are about 150 types of HPV; most clear the body but about 40 can linger and cause various types of cancer, which, in women, primarily is of the cervix.
- Randomized controlled clinical trials have shown that a vaccine affords nearly 100% immunity to infection by four of the most common types of HPV, which cause 70% of all cervical cancer—a disease that kills on average more than one Canadian woman per day and leaves survivors infertile—and 90% of genital warts.
- The trials were conducted on 16–23-year-old females.

- The tests show the immunity to last at least 5 years, but it will take decades to learn how long the immunity lasts.
- The vaccine works only if administered before exposure to the viruses.
- In 2006 in Canada: about 1,350 women were diagnosed with cervical cancer and 390 died, more than 22,000 were diagnosed with breast cancer, and the number who died from cervical cancer is a small fraction of those who died from lung and ovarian cancer.

So, the medical research is very clear. The degree of immunity afforded by the vaccine is outstanding. Should the course of action be obvious? Based upon medical facts, certain recommendations and actions have been taken:

- Medical experts advise giving the vaccine before girls are sexually active.
- 12–14-year-old girls were offered the drug at public expense and with parental consent in four Canadian provinces in the fall of 2007.

What normative considerations led the medical experts to their recommendation? On what normative basis did the four provinces make their decisions? On the one hand, it might be argued that the trials were so positive that it would be unethical not to make the vaccine available. On the other hand, one might wonder why and how different provinces with the same facts reached different conclusions.

These considerations lead us to the question of how the research can be applied in specific cases. What ought particular parents to do: provide or withhold their permission? On what basis should they decide? We will argue that there can be great variability of application, many bases for decision, and that the same theory or knowledge can be applied legitimately in different ways according to the context and the situational appraisal. Considerations of the child's size and physical and emotional maturity, the history of cervical cancer in the child's lineage, and the existence of developmental disorders might all come into play. Consider the following sketches of arguments by a number of parents reported in the newspaper article.

Parent 1: Anything that will protect my daughter from cancer is worth the risk. (Spoken as justification for giving permission for inoculation)

Parent 2: I have fear of side effects, question the motives of the drug company, and feel queasy about dosing girls as young as 10 with protection against an STD. (Spoken as justification for declining permission for inoculation)

Parent 3: At this age, kids are pretty innocent and this is not something they should have to worry about. (Spoken as justification for declining permission)

Parent 4: If a doctor said I can provide a vaccine against cancer of the lung, I wouldn't think twice about it. (Spoken as justification for giving permission)

Parent 5: It's not like vaccinating your kid against polio. There is no epidemic of cervical cancer. (Spoken as justification for declining permission)

Parent 6: I'll wait and see whether more is known in a couple of years' time. (Spoken as justification for declining permission)

Parent 7: The vaccine may promote early sexual behavior or unsafe sex or a belief that it is ok to be sexually active. (Spoken as justification for declining permission)

Parent 8: We can't trust the medical community to know what is best for our children. (Spoken as justification for declining permission)

All of these parents knew that the research evidence by itself was insufficient as a guide to action. Other normative considerations—some political, some religious, some ethical, some prudential, some pragmatic—had to be brought into play before a decision could be made on whether to give or decline permission for their daughters to be vaccinated. Not only is there room for judgment, it is demanded, because the medical evidence by itself does not imply which action to take. Even though different decisions were reached, all of the parents were using the same research-based knowledge. Moreover, we find it rhetorically striking that, even though the medical evidence was overwhelmingly positive about the effectiveness of the vaccine, 75% of the small sample of parents withheld their permission.

In education, the empirical evidence on effectiveness of treatments is never so clear. We must be cautious, however, not to view the link between knowledge and action as weaker than it is. Although theories and knowledge do not prescribe precisely, they can constrain action. Some actions are not in accord with the medical research on the vaccine against cervical cancer. For example, a program of vaccination in senior secondary school would not make sense because of the increased chance of exposure to the viruses through intercourse.

27.3.4 Role of Values and Choice in Application

We saw in the previous section that application involves normative considerations about whether and how to apply a given theory. Ultimately, we must have a clear conception of what we to try to achieve when we are applying knowledge because what is desired can alter the auxiliary hypotheses needed for application. For example, if we want only a rough prediction or want to make only a small modification in the world, then we might take into account through auxiliary hypotheses only some of the more important influences that lie outside the scope of the theory. If a parent wanted above all else to reduce the risk of cervical cancer in a daughter to a minimum (perhaps because of a personal traumatic experience with the cancer), then the parent might not be satisfied with the low risk that exists even without vaccination and opt for the inoculation in order to achieve the lowest risk possible.

If, however, a parent has other beliefs, such as that drugs are inherently dangerous and that the risks of catching the disease (known to be low) are not clearly higher than the unknown risks from drug side effects, the parent could easily justify foregoing the vaccination. We wish to make clear that we do not advocate an anything-goes policy. Take, for instance, the parents who based a decision to permit their daughter's vaccination on the grounds that anything that protects her from cancer is worth the risk. It is unlikely the parents actually believe this justification. For example, one way to protect a child from skin cancer due to sun exposure is never to permit the child to go outdoors. All foods, even organic ones, expose the body to some carcinogenic substances. A way to avoid exposure is not to eat, which is a ridiculous course of action. So, the parent does not mean *anything* that protects against cancer is worth the risk because some things that protect against cancer

impose an even more severe risk of other, even more undesirable consequences. All action requires a trade-off between competing values. It is often difficult in such trade-offs to see one value trumping all others.

27.3.5 *Summary*

We have attempted to portray the main points of this section in Fig. 27.2. First, note that theory is related to some phenomena through the relationship of explanation—the theory explains the phenomena. The relationship between the theory and some application situation might also be one of explanation; but it might also be one of prediction, description, intervention, or perhaps other possibilities. We have used the overarching expression *applies to* to capture all these possibilities. The figure shows that the application emanates from the theory with two supplements indicated by the addition symbols, first, auxiliary hypotheses about the workings of the application situation and, second, normative considerations for and against the application. The point is to show that application is not a direct line from theory to the application situation.

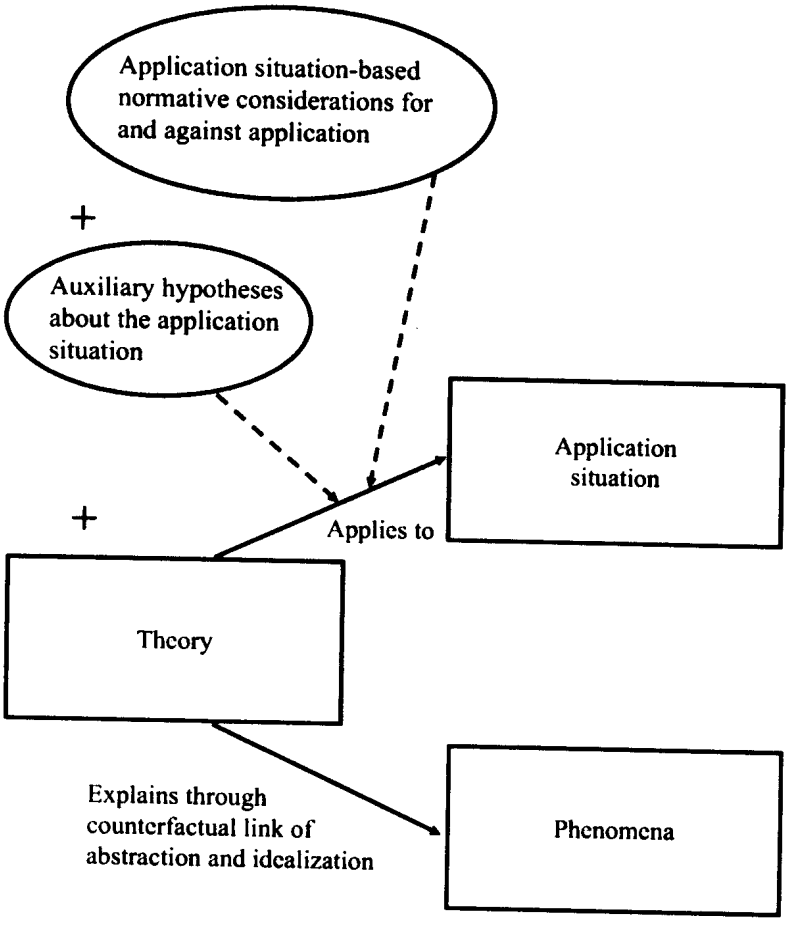

Fig. 27.2 The relationship among theory, phenomena, and application situations

27.4 An Intervention Study

Now that we have outlined our theoretical machinery, we turn to a longitudinal, early literacy intervention study that two of us helped to conduct (Phillips, Norris, & Mason, 1996). We introduce this example because it matches to a large degree the type of study for establishing what works that falls under the Gold Standard and because we have full information on the study, including access to the raw data. There were three treatment groups—a school-only treatment, a home-only treatment, and a home and school treatment—and a control group. Treatment children were given extra instruction in early literacy concepts using a series of *Little Books* (McCormick & Mason, 1990) in Kindergarten and the effects were followed until the end of Grade 4. Positive effects were strongest at the end of Grade 2 and in this order: school treatment, home and school treatment, home treatment.

In the school-only treatment, children and teachers read *Little Books* in school in addition to the approved language arts program. In the home-only treatment, children and their parents read the same *Little Books* at home, and the children received the approved language arts program in school. In the home and school treatment, *Little Books* were read in both settings in addition to the approved language arts program. In the control group, *Little Books* were not used at all and the children received the approved language arts program.

The study randomized on classroom and analyzed data by students, thus not keeping constant the unit of analysis. Covariance analysis was used to remove the effects of preexisting differences among the groups. As such, it was not a strictly Gold Standard study but met well the criteria of "an intervention backed by 'possible' evidence of effectiveness" (US ED, 2003, p. 11). However, every point made in the subsequent discussion would apply even if the study had met strictly the Gold Standard.

Figure 27.3 presents a scatter plot of the relationship for the school-only treatment children between their pretest scores at the beginning of Kindergarten (Metropolitan Reading Readiness Test [MET 1], Nurss & McGauvran, 1987) and their posttest scores at the end of Grade 2 (National Achievement Test [NAT II], Wick, Fraenkel, Mason, Stewart, & Wallen, 1989). Each point represents a single child's pair of scores. Scores are in standard deviation units so that scores of zero are average for both measures. The diagonal line represents children whose relative standing on both measures is the same, that is, their scores are above or below the mean by the same number of standard deviation units on each test. To the right and below the diagonal, children had a higher relative standing on the pretest than they did on the posttest. Pick any of those points, and you will see that the child represented by that point has a higher standard deviation score on MET I than on NAT II. To the left and above the diagonal, the relative standing of children was higher on the posttest than on the pretest.

The intersection of the vertical and horizontal lines is the centroid for the control group, that is, the average scores in standard deviation units that the control group children received on both measures. You can see immediately that on average the

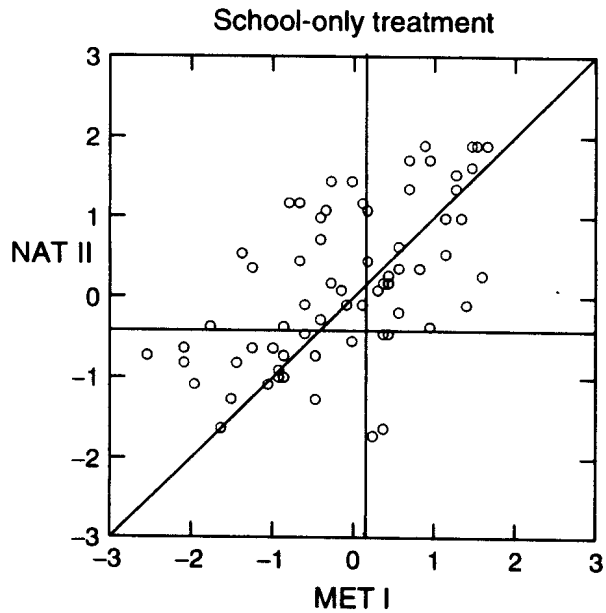

Fig. 27.3 Scatter plot for the school-only treatment children between their pretest scores at the beginning of kindergarten (MET I) and their posttest scores at the end of second grade (NAT II)

control group children performed relatively better on the pretest than they did on the posttest because the centroid is to the right and below the diagonal. Students whose scores fell to the right of the vertical line did better than the control group on the pretest; to the left, they did worse than the control group on the pretest. Children above the horizontal line did better than the control group on the posttest; below the line, they did worse than the control group. We also can combine the information from looking at those falling left and right of the vertical line and those falling above and below the horizontal line: children in the lower-left quadrant did worse on both measures than the control group children; those in the upper-right did better on both measures than the control group; children in the lower-right lost ground compared to the control group because they were above the average of the control on the pretest measure but below the average on the posttest; in the upper-left quadrant, children gained ground because they scored lower than the average for the control children on the pretest but higher than average on the posttest.

Look now to Fig. 27.4, which provides the same information for the home and school treatment children. The data are distributed differently. In particular examining the lower-right and upper-left quadrants, you can see that, compared to the school-only treatment, a greater proportion of the home and school treatment children lost ground with respect to the control group (0.10 versus 0.06) and a smaller proportion gained ground (0.14 versus 0.25). So, the home and school treatment did not work as well as the school-only treatment.

Examine Fig. 27.5 for the home-only treatment children in which the contrasts to Fig. 27.3 are even starker than for Fig. 27.4. Compared to the home and school

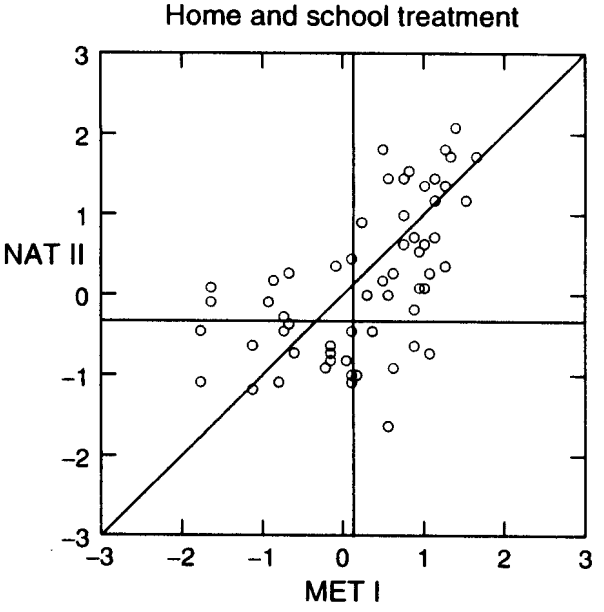

Fig. 27.4 Scatter plot for the home and school treatment children between their pretest scores at the beginning of kindergarten (MET I) and their posttest scores at the end of second grade (NAT II)

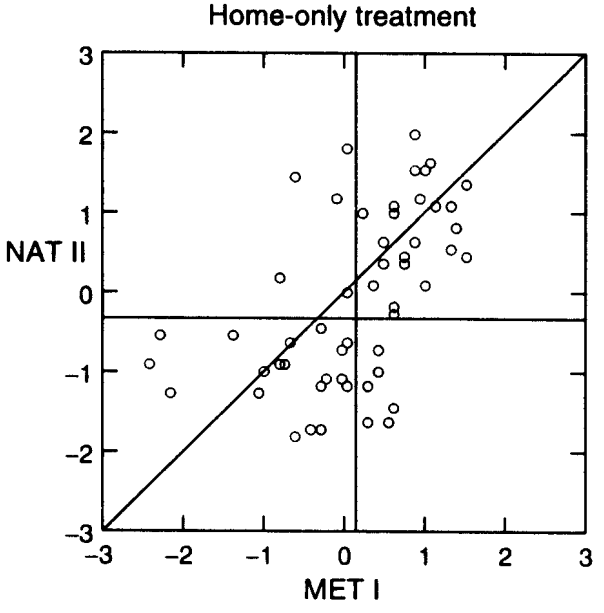

Fig. 27.5 Scatter plot for the home-only treatment children between their pretest scores at the beginning of kindergarten (MET I) and their posttest scores at the end of second grade (NAT II)

treatment, an even greater proportion lost ground with respect to the control group (0.11) and an even lower proportion gained ground (0.08).

We wish now to use this example to motivate a general analysis of the desired conclusion from an intervention study. In general, we wish to infer from a claim of the form '*a* caused *b*' (i.e., what happened in a particular case) to a claim of the

form '*A*s cause *B*s' (i.e., what happens in general) or from '*a* did not cause *b*' to '*A*s do not cause *B*s'. The first type of claim, specific causal claims, includes past tense singular claims about what has happened. The second type of claim, general causal claims, contains tense-less general claims about standing states or conditions. There is often an implied *usually, generally, frequently* as part of claims in this latter category. The distinction is very similar to that drawn by Campbell and Stanley (1963, p. 5) between internal validity ("Did in fact the experimental treatments make a difference in this specific instance?") and external validity ("To what populations, settings, treatment variables, and measurement variables can this effect be generalized?").

Although it is conceivable, perhaps even likely, that some results may be particular to individual experiments, the point of randomized experimental design is lost if one cannot expect some level of generality from the research. We are always interested in moving from '*this intervention* caused an effect of size ε in *this sample*' to '*interventions of this type* cause effects of size ε in *samples from this population*.' This is the type of information we wish to gather from Gold Standard research.

We have chosen to use explicitly causal language, rather than Campbell and Stanley's "make a difference" (1963, p. 5). Sometimes there is an attempt to avoid the imputation of causation on the grounds that it implies mechanical or deterministic systems. Most such attempts are unsuccessful (see, e.g., Ennis, 1982). We say unsuccessful because causation is implied by such language as *brought about, led to, succeeded in creating, made a difference*. We believe also that the concept of intervention contains a causal implication. Interventions are actions we take with specific intentions to alter the course of events from what they would have been otherwise. Nevertheless, when we look at what happens to individual students, some might experience an effect equal to ε, some experience an effect larger than ε, some experience an effect smaller than ε, some experience a negative effect, and some experience no effect at all. Results like this typically occur. The Gold Standard is not about what happens to individuals but about what happens to the group on average.

So, even with Gold Standard evidence, there is still a decision about what to do on the basis of it: Are the gains by those who gain worth the losses by those who lose? Is there an intervention with more acceptable trade-offs? Is the monetary cost worth the gains that are found? Therefore, based upon this study, what should schools do? First of all, it is not immediately obvious that they should adopt the *Little Books* intervention. There is an effect size of about 2 standard deviations needed to bring the children targeted by the intervention up to the mean of their peer group. The *Little Books* interventions produced an effect equal to about 0.25 of a standard deviation. Not all children profited from the intervention, and some even fell behind. Unfortunately, such is typically the case even with interventions that pass the Gold Standard of effectiveness. In education, there is rarely one approach that works for everyone. Perhaps the schools would like to wait to find an intervention that works for more children. Perhaps they would like to try a combination of interventions. Of course, combining interventions can lead to complications because a positive effect that an intervention has when used in

isolation may not be sustained when used in combination with something else. Basically, there is so much left to decide, even when the evidence is in and it points to effectiveness!

27.5 Conclusions and Policy Implications

We hope to have shown that evidence does not determine action, in the sense of leaving open one and only one possible way to proceed. The evidence alone—even from Gold Standard research—cannot tell us what to do. Even with clear conclusions, much is left to individuals with situation-specific knowledge to decide how research is best applied in their contexts. This is the same point that we made earlier working from a theoretical perspective on the nature of theories and research-based knowledge.

Similar points are argued in a very insightful set of chapters in a book edited by Thomas and Pring (2004). For example, Cordingley (2004) made the important and often overlooked point that, even in the context of full evidence (if such can be imagined), there will always be a role for professional judgment in deciding how the evidence is best applied to particular settings and particular students. Eraut (2004) made a similar point to Cordingley's in the context of medicine, namely, that the idea of evidence-based practice seems to presuppose the incorrect view that somehow evidence can determine what ought to be done in practice. Eraut argued that, in addition to research-based evidence, practitioners need to draw upon knowledge derived from their own experience, which he calls practice-based evidence. The main point of Hodkinson and Smith's (2004) chapter is that there is no such thing as safe research, research that points with perfect reliability to a course of action. Above all, they claimed, the relationship between research and practice is imbued with an uneliminable political element.

An important caution arising from our analysis is that policy makers need to be fully aware of the politics involved in the use of educational research. We believe they need to know and grasp the significance of at least the following: that, in using the results of scientific educational research to guide practice, even evidence based upon Gold Standard research cannot by itself determine decisions about what to do; that the use of scientific results involves a mediation between abstract and general scientific knowledge and concrete and specific situational knowledge; that they and other educational practitioners are the mediators; and that mediations are rarely clear-cut, because the same knowledge can be applied in different ways in different contexts and at different times in the same context.

Perhaps education programs, particularly those aimed at educational administrators at the graduate level, could focus upon some key abilities needed to use research-based educational knowledge. The ability to formulate reasonable auxiliary hypotheses to mediate between theories and concrete educational situations is unlikely to be something that comes naturally, even to individuals who realize that such hypotheses are needed. Likewise, the ability to employ normative considera-

tions in conjunction with the best scientific and situational knowledge is not something that currently is given much attention by faculties of education.

How is it best to teach these abilities? At least, we conjecture, through plenty of practice with examples that: demonstrate the pitfalls that arise when it is assumed that scientific knowledge is directly applicable; demonstrate the variety of auxiliary hypotheses and normative principles that must be brought to bear for successful application; encourage the explicit formulation of auxiliary hypotheses and normative principles through situational appraisals; encourage the conjecture, consideration, and evaluation of alternative application routes for a given theory in a given context that depend upon different desired outcomes of the application; and encourage the evaluation of whether applications are consistent or not with the theory being applied.

If we wish scientific educational research to serve the public good by providing part of the basis for many of our educational practices, then scholars of education have a role in showing policy makers how they can use scientific educational research results in their practice and in providing policy makers the opportunity to acquire the knowledge, skills, and dispositions to use science wisely. The IES could do well by including an extensive elaboration in their documentation of the role of evidence in implementing changes in educational practice. Gold Standard research is next to impossible to conduct in authentic educational settings and, where it is possible, provides no absolute guidance on what to do. Finally, this last conclusion is not meant to imply that Gold Standard research is not important. Quite the contrary—it is important to have the most robust evidence possible for making educational policy. The conclusion is meant to reiterate that evidence based upon Gold Standard research just does not have the degree of authority that many advocates proclaim it has.

References

Anderssen, E., & Alphonso, C. (2007). Should your daughter get the needle? *The Globe and Mail, Alberta edition*, A1, A16–17.

Bush, G. W. (2002, November 5). *Statement on signing legislation to provide for improvement of federal education research, statistics, evaluation, information, and dissemination, and for other purposes*. Retrieved February 14, 2008, from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname = 2002_presidential_documents&docid = pd11no02_txt-21.pdf

Campbell Collaboration. (n.d.). *Education coordinating group Homepage*. Retrieved June 28, 2008, from http://www.campbellcollaboration.org/ECG/index.asp

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of research in science teaching* (pp. 171–246). Chicago: Rand McNally.

Cochrane Collaboration. (n.d.). *Homepage*. Retrieved October 3, 2007, from http://cochrane.org/docs/descrip.htm

Code, L. (1987). *Epistemic responsibility*. Hanover, NH: Brown University Press.

Cordingley, P. (2004). Teachers using evidence: Using what we know about teaching and learning to reconceptualise evidence-based practice. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 77–87). Maidenhead, Berkshire, UK: Open University Press.

Dewey, J. (1929/1984). The quest for certainty: A study of the relation of knowledge and action. In J. A. Boydston (Ed.), *John Dewey, The Later Works, 1925–1953* (Vol. 4: 1929). Carbondale, IL: Southern Illinois University Press. (Original work published 1929)

Dewey, J. (1929–30/1984). The sources of a science of education. In J. A. Boydston & K. E. Poulos (Eds.), *John Dewey, The Later Works, 1925–1953* (Vol. 5: 1929–1930). Carbondale, IL: Southern Illinois University Press. (Original work published 1929–30)

Ennis, R. H. (1982). Research notes: Abandon causality? *Educational Researcher, 11*(7), 25–27.

Eraut, M. (2004). Practice-based evidence. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 91–101). Maidenhead, Berkshire, UK: Open University Press.

Education Sciences Reform Act of 2002. Pub. L. No. 107–279, 116 Stat. 1940. (2002).

Hodkinson, P., & Smith, J. K. (2004). The relationship between research, policy and practice. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 150–163). Maidenhead, Berkshire, UK: Open University Press.

Hume, D. (1739–40/1962). *A treatise of human nature: Being an attempt to introduce the experimental method of reasoning into moral subjects.* (Edited with an introduction by D. G. C. Macnabb). Cleveland, OH: World Publishing. (Original work published 1739–40)

McCormick, C. E., & Mason, J. M. (1990). *Little books.* Glenview, IL: Scott, Foresman.

No Child Left Behind Act of 2001. Pub. L. No. 107–110, 115 Stat. 1425. (2002).

Norris, S. P. (2000). The pale of consideration when seeking sources of teaching expertise. *American Journal of Education, 108*(3), 167–195.

Norris, S. P., & Kvernbekk, T. (1997). The application of science education theories. *Journal of Research in Science Teaching, 34*(10), 977–1005.

Nurss, J. R., & McGauvran, M. E. (1987). *Metropolitan reading readiness test.* San Antonio, TX: The Psychological Corporation.

Phillips, L. M., Norris, S. P., & Mason, J. M. (1996). Longitudinal effects of early literacy concepts on reading achievement: A kindergarten intervention and five-year follow-up. *Journal of Literacy Research, 28*(1), 173–195.

Rudner, R. (1953). The scientist *qua* scientist makes value judgments. *Philosophy of Science, 20*(1), 1–6.

Suppe, F. (1977). *The structure of scientific theories* (2nd edn.). Urbana, IL: University of Illinois Press.

Suppe, F. (1989). *The semantic conception of theories and scientific realism.* Urbana, IL: University of Illinois Press.

Thomas, G. (2004). Introduction: Evidence and practice. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 1–18). Maidenhead, Berkshire, UK: Open University Press.

Thomas, G., & Pring, R. (Eds.). (2004). *Evidence-based practice in education.* Maidenhead, Berkshire, UK: Open University Press.

United States Department of Education. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide.* Washington, DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Available from http://ies.ed.gov/ncee/pubs/evidence_based/evidence_based.asp

Wick, J. W., Fraenkel, J. R., Mason, J. M., Stewart, J., & Wallen, N. E. (1989). *National achievement test.* Chicago: American Testronics.