The Good, the Bad, and the Ugly of Evidence-Based Practice

Marcel Bouffard
University of Alberta, Canada

Greg Reid
McGill University, Canada

The evidence-based practice (EBP) movement has been extremely influential over the last 20 years. Fields like medicine, physiotherapy, occupational therapy, nursing, psychology, and education have adopted the idea that policy makers and practitioners should use interventions that have demonstrated efficiency and effectiveness. This apparently straightforward idea is beginning to affect adapted physical activity; however, researchers and practitioners in our field often appear to be unaware of fundamental questions related to them. The major purpose of this paper is to outline and discuss 10 of these fundamental questions. This analysis leads us to conclude that EBP is a good direction to pursue in adapted physical activity if we develop a type of EBP congruent with the main tenets of our field.

The idea that practitioners ought to deliver services based on the best evidence available is intuitively appealing and has gained widespread acceptance (Worrall, 2002). It is a common sense idea that is typically viewed favorably by researchers, funders, practitioners, service users, and policy makers. To request that, whenever feasible, service providers use the best research evidence about the likely benefits of an intervention is a sensible request that can be justified on both ethical and financial grounds.

A common goal of professions like adapted physical activity (APA), medicine, physiotherapy, nursing, psychology, education, and others is to deliver treatments or interventions that produce the intended outcome without negative consequences. Advocates of Evidence-Based Practice (EBP) share the same concern. They want to know if the treatment positively changes the target behavior, and they request evidence that the treatment does not produce significant harm.

In this paper, the expression Evidence-Based Medicine (EBM) will be used to refer to practices related to this movement in medicine. Other labels like “evidence-based healthcare,” “evidence-based education,” and “empirically supported treatments” have been used to describe similar movements in other professions or...
disciplines. We will use EBP to refer to all these movements, including EBM. The inclusion of EBM under EBP is justified given that EBM is the parent discipline of EBP.

The literature about EBP is abundant. A search conducted on MEDLINE during the month of June 2011, using the expressions “evidence-based medicine” produced more than 42,000 hits for both expressions. A similar search on Google using the expressions “evidence-based medicine” and “evidence-based practice” produced more than 9 million hits. Despite the huge literature, it is possible to identify a number of core ideas frequently associated with EBP. A major purpose of this paper is to outline these ideas and raise some fundamental questions related to them.

**Brief History of Evidence-Based Practice**

The use of evidence to support claims has had a long history. It is believed that Aristotle (384–322 BC) was the first person to “really think out the problem of evidence” (Hooker, 1996). The more recent origins of EBP are found in medicine. In the 1830s, the French physician Pierre Charles Alexandre Louis introduced a movement called *Médecine d’observation*, which essentially stated that physicians “should not rely on speculation and theory about causes of disease, nor on single experiences, but they should make large series of observations and derive numerical summaries from which real truth about the actual treatment of patients will emerge” (Vandenbroucke, 1996, p. 1335). During the 19th Century, *la médecine d’observation* was ultimately rejected by physicians who contended that “medicine was about individual patients and not about groups” (Vanderbroucke, 1996, p. 1137). In 1972, Archie Cochrane published his influential *Effectiveness and Efficiency* book in which he recommended standards for medical research. Briefly, he argued that the randomized control trial (RCT) was the best way to provide evidence and he judged physicians’ opinions to be of low value. *Clinical Epidemiology* (Sackett, Haynes, & Tugwell, 1985), which announced that epidemiology, a discipline traditionally concerned with population health issues, could be applied to the care of patients preceded the work of the Evidence-Based Medicine Working Group (EBMWG) at McMaster University in Canada (EBMWG, 1992). This work has been extremely influential in medicine and other disciplines (e.g., nursing, occupational therapy, physiotherapy, psychology, education, social work).

The apparently simple idea of making decisions based on research evidence has led to a paradigm shift according to advocates of the EBMWG: “A new paradigm for medical practice is emerging. Evidence-based medicine deemphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research” (EBMWG, 1992, p. 2420). Currently, the EBP movement is supported by institutions such as Evidence-Based Practice Centers, publishers of textbooks and journals, as well as institutions offering courses and workshops (e.g., the Cochrane Collaboration, the Agency for Healthcare Research and Quality, the What Works Clearinghouse). Evidence-based medicine is a general philosophy of medicine, defining acceptable questions and appropriate evidence (Solomon, 2009). Over the last two decades, the EBP movement has affected the prioritization of research questions, research methods, research funding, and service delivery. It is sometimes claimed that the movement functions as a regulatory system (Lambert,
Evidence-Based Practice and Adapted Physical Activity

In our field, supportive arguments about the potential of EBP in APA have been advanced. Recent publications (Hutzler, 2006, 2010, 2011; Jin & Yun, 2010) and submissions to this journal indicate a high level of enthusiasm about EBP. For example, Jin and Yun (2010) state that “teaching practices supported by the best available evidence hold the greatest hope [italics added] of meeting students’ unique needs” (p. 50). These documents suggest that advocates of EBP in our field have embraced ideas from other fields.

We submit that careful attention to some EBP ideas is essential to develop an EBP tradition that is congruent with the major tenets of our field. In this paper, we ask some fundamental questions about EBP such that we recognize the potential pitfalls of this movement and avoid them. The EBP movement has evolved quickly over the last 20 years. We are now in a better position to understand some of the problems, challenges, and issues that have been faced by other fields, and we can now use this knowledge to build a better future in APA.

The purpose of this paper is to introduce the reader to some of the issues related to the apparently straightforward idea of using best research evidence to justify practice. Issues and dilemmas recur in the vast literature about EBP. We ask 10 questions related to EBP that attempt to summarize some of the concerns we have about this movement. The 10 questions should function as a preliminary overview of EBP before readers delve more carefully into specific details. We present these questions because oftentimes they do not seem to be a part of current debates in APA. By asking them, we hope that the reader will better appreciate some of the complexities and challenges of this movement. We think that our field is at risk for developing EBP without giving sufficient attention to fundamental questions. We hope that our presentation will highlight some critical issues related to EBP such that some of its potential pitfalls could be alleviated.

Although we have concerns about some understandings of EBP, we think that careful attention to the 10 questions raised may lead to better practices in APA. We believe that unfettered and professional debate is essential to the evaluation of ideas. To question the foundations of a movement or a practice is not necessarily to “deny its value, but rather to stimulate a judicious and balanced appraisal of its merits” (Ashcroft & ter Meulen, 2004, p. 119). We think that “one can never know the true worth of an idea unless one is free to examine it” (Cole, 2005, p. 6).

Ten Questions

Question 1: What is Evidence-Based Practice?

There is no consensus definition of EBP (for a good sample of definitions see “Definitions of Evidence Based Practice”, n.d.). At this moment, seeking absolute consensus about what EBP means, or should mean, is not possible. The expression
“EBP” is an umbrella term. As noted by Wittgenstein (1953), we often use the same name to talk about members of a heterogeneous category. Although the expression EBP is often used, we can observe diversity in the details of the different understandings advanced. Any precise definition of EBP will fail to account for all cases where the expression EBP is used. The lack of a consensus definition of EBP is not surprising. As noted earlier, EBM has impacted other disciplines and professions that have had different goals. The various understandings of EBP reflect the particular features of different professions. Further, in reaction to critiques over the last 20 years, proposed definitions and understanding have changed over time.

Despite the different understandings of EBP, a central concern is to establish “what works?” This concern can be illustrated by the following quotations made by EBP promoters:

1. “The idea is not to establish the ‘truth’ but rather to find out ‘what works’” (Holmes, 2006, p. 188).
2. “Simply stated, evidence-based treatment is the use of treatment methodologies for which there is scientifically collected evidence that the treatment works” (Hayes, 2005, p. 1).
3. “Research should provide decisive and conclusive evidence that if teachers do X rather than Y in their professional practice, there will be a significant and enduring improvement in outcome” (Hargreaves, 1996, p. 53).
4. “EBP occupies the moral high grounds because its practitioners do ‘what works’” (Tanenbaum, 2005, p. 169).

In addition to the emphasis on what works, we think it is possible to provide three definitions that capture some frequently encountered understandings of EBP. First, EBP sometimes refers to the use of evidence to guide practice. We find this definition too general to be useful for the purpose of this manuscript. Second, EBP refers to the use of best research evidence to guide practice. People using this definition emphasize knowledge production, synthesis, dissemination, and use. Third, it is often advanced that EBP integrate best research evidence, practitioner expertise, and participant preferences, values, and goals in its definition (e.g., see Spring, 2007; Straus, Richardson, Glasziou, & Haynes, 2005). According to the third definition, all these factors must be considered to determine the best course of action for a person. Most of the questions asked in this manuscript assume the second definition of EBP. This particular focus is justified given that issues related to knowledge production and synthesis must be addressed before we can address problems related to practitioner expertise and participant preferences, values, and goals.

It appears useful to differentiate between three major types of questions related to EBP. First, epistemological questions are prominent in the discourse surrounding EBP. People engaged in these questions typically ask “What is evidence?” “How do we know?” “Why should we give more credence to claim A than to claim B?” “Is one method of knowing really better than another one?” The second type of questions focus on problems related to knowledge synthesis, translation, dissemination, and use. The fundamental concern of the two types of questions relate to efficiency and effectiveness (Cochrane, 1972). Efficiency is concerned with whether a treatment works; hence internal validity (Shadish, Cook, & Campbell, 2002) is
Evidence-Based Practice

of central importance in an efficiency study. If treatment efficiency can be demonstrated, usually under tightly controlled conditions without significant negative side effects, then effectiveness studies designed to know whether the treatment works in a “real-world” situation can be conducted. Effectiveness is akin to external validity (Shadish et al., 2002). Finally, sociopolitical questions about power and control (Foucault, 1980) are the third type of questions.

To some, it may appear disconcerting that there is no consensus definition of EBP. Instead of conceiving EBP as a set of homogeneous practices, it is better to think of it as a set of heterogeneous practices. Different professions have developed their own EBP, and like these professions, we submit that we have the opportunity to develop a set of practices that are congruent with the major aspirations of adapted physical activity (e.g., empowerment, self-determination, philosophy of individuality; Reid, Bouffard, & MacDonald, 2011). We can use the knowledge gained in other fields to develop our own EBP.

Question 2: What Counts as Evidence?

In a broad sense, evidence usually refers to any information bearing on the truth or falsity of a proposition; however, both in philosophy and in EBP, there has been a debate about what sort of things are eligible to serve as evidence. Although some have claimed that “any empirical observation about the apparent relationship between events constitutes potential evidence” (Guyatt & Rennie, 2002, p. 6), numerous EBM advocates contend that evidence should be derived from “certain types of research” (Haynes, 2004, p. 234) with the RCT as the “gold standard” of research and the best source of evidence for decision making. Good meta-analyses based on RCT have also received high regard. Of course this implies that other research strategies like quasi-experimental designs, observational studies, or qualitative studies have low evidential status, if any. In addition, traditional medical knowledge based on biology, biochemistry, physiology, and pathology is viewed as untrustworthy. In other words, EBM proponents have an epistemological theory stating that empirical evidence, preferably obtained using RCT, have a greater epistemological value than knowledge based on theories of biology, biochemistry, physiology, pathology, etc. To them theoretical knowledge has often mislead practice and it must be replaced, or at least supplemented, by empirical evidence.

A statement on the back cover of the User’s Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice is indicative of the kind of epistemology guiding EBM: “We wish that this book, instead of a biochemistry textbook, was a required reading for medical students. We hope that all clinicians will not only read, but study this reference because we believe that patients would certainly benefit” (Guyatt & Rennie, 2002). Stated differently, the results of empirical investigations, preferably obtained using RCT, are epistemically superior (i.e., provide more reliable knowledge) to theorization or explanation produced by basic research. “Evidence-based medicine de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision-making and stresses the examination of evidence from clinical research” (EBMWG, 1992, p. 2420). In summary, EBP is guided by a pragmatist viewpoint where “the function of inquiry is . . . to enable us to act more effectively” (Rorty, 2005, p. 838). Empirical inquiry, preferably using RCT, leads to better decisions and interventions.
A central concern of EBM proponents is to determine whether causal assertions are warranted and, to them, the RCT is the gold standard. Advocates of EBM have adopted a manipulative, successive view of causality (Cook & Campbell, 1994). That is, a cause must be manipulable and precede an effect. They prioritize causal descriptions over causal explanations (see Shadish et al., 2002, pp. 9–12). Causal descriptions are often made by researchers who use group experimental designs. In this case, the intent of the researcher is to determine the impact of manipulating A on B. If A precedes B, and there is covariation and plausible alternative hypotheses are ruled out, then the researcher can safely conclude that A causes B. It is essential to note that when a causal description is made, the reasons or the mechanisms relating a change in A to a change in B typically remain unspecified. A researcher who intends to produce a causal explanation will attempt to clarify “the mechanisms through which and the conditions under which that causal relationship holds” (Shadish et al., 2002, p. 9). In a classical treatise on experimental and quasi-experimental designs, Shadish et al. (2002) state that “[t]he purpose of experiments is not to completely explain some phenomenon; it is to identify whether a particular variable or small set of variables makes a marginal difference in some outcome over and above all the other forces affecting that outcome” (p. 457). Evidence-based medicine advocates accept that science is concerned with the discovery of causes but they are reluctant to prioritize the search for causal mechanisms as typically outlined in theories. To them, theorization has often led to inappropriate interventions and good empirical studies are a better guide to interventions.

Lurking behind descriptive and explanatory views of causality are some research assumptions (Slife, 1998). The question “What counts as evidence?” receives different answers based on the assumptions the researcher is making. Advocates of EBM have endorsed a successive empiricist understanding of causality, while proponents of realism argue in favor of explanatory theories. Pragmatists are likely to argue that what matters is what works, not how it works. On the other hand, realists (e.g., Thompson, 2010) argue for the necessity of having a credible explanatory model, or theory, to explain relationships between presumed cause(s) and effect(s). Realists typically argue that we have no evidence until we have a credible explanatory theory.

The different understandings of causality and the philosophical viewpoints behind them have implications for the development of EBP in APA. There is no doubt that research using RCT has produced significant and important understandings; however, we find the causal description viewpoint to be limitative. It should be supplemented by an attempt to understand the mechanisms relating independent variables to dependent variables (for elaboration see Kazdin, 2007; Kraemer, Wilson, Fairburn, & Agras, 2002). There are two major viewpoints about causality and we should welcome both viewpoints for their potential contribution to our field.

**Question 3: What Are the Rationales Behind the Different Hierarchies of Evidence?**

A number of hierarchies of evidences have been offered. For example, the Agency for Healthcare Research and Quality (2002) reviewed 106 systems (for other examples see Table 1 in Hutzler, 2011, and Table 1 in Jin & Yun, 2010). Many more hierarchies have been proposed since 2002. The main intent behind these
hierarchies, also called levels, has been to rank order research methods from best to worst. A quick perusal of these hierarchies reveals no perfect agreement across the different hierarchies, but a general consensus emerges when these hierarchies are compared. Randomized clinical trials, preferably conducted with large sample sizes and/or the meta-analysis of RCT, are typically at the top of the hierarchy. At the lower level, observational studies, qualitative studies, clinical knowledge, and theoretical knowledge as offered by disciplines like biology, biochemistry, physiology, or basic mechanisms of disease are viewed as producing inferior knowledge. Stated differently, producers of hierarchies have advanced epistemic positions stating when knowledge claims are warranted or less warranted.

The most disconcerting aspect about these hierarchies is that the rationale behind them is never presented in detail. At best, we are told that some experts discussed different research methods that they then ranked or graded. This is ironic given that advocates of EBP typically view expert opinion as one of the lowest form of evidence (Goldenberg, 2009). As noted by Loughlin (2006),

the campaign to restrict the meaning of the term “evidence” to certain sources of knowledge has been a campaign to effectively declare other sources of knowledge either illegitimate or at least inferior. This is how authors can come to assert the existence of epistemic “hierarchies” in the absence of any sound arguments explaining why we should regard some sources of knowledge as inherently superior to others. (p. 289)

The lack of a clear rationale for the hierarchies impairs the evaluation of ideas. This ignorance of fundamental knowledge issues is a strange state of affairs. How can we engage in unfettered debate if basic assumptions are not stated?

To develop hierarchies of evidence is to fundamentally engage in applied epistemology. Clearly, different epistemologies underlie different research traditions (Alvesson & Sköldberg, 2009; Cook & Campbell, 1979; Crotty, 1998; Slife 1998), and the production of hierarchies may lead to the rejection of certain research traditions. The development of hierarchies within our own field may create profound divisions in APA. We need to develop EBP that are sensitive to different worldviews and methodologies. Instead of asking which source of knowledge is universally better, and ranking them, it might be preferable to ask how each source might complement each other.

**Question 4: How Do We Generalize From Aggregate to Person?**

In most hierarchies of evidences, the RCT has been viewed as the best research strategy to determine the magnitude of the treatment effect and justify the causal description inference (Raudenbush, 2005). If EBM is “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual [italics added] patients” (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, p. 71), then the results of RCT could lead to incorrect inference. As noted earlier in this journal when discussing issues related to aggregation (Bouffard, 1993), typical strategies used to analyze data from RCT justify inferences from sample aggregate to population aggregate. These strategies do not justify
generalizations from population aggregate to an individual person. Bouffard (1993, Table 1) reviewed a number of studies supporting this claim. More recent physical activity studies, labeled “The Heritage Project,” reported similar findings (Bouchard & Rankinen, 2001). Briefly, knowing that, on average, treatment X is better than treatment Y for condition C does not imply that treatment X will work for a particular person with condition C (Upshur, 2005).

Meta-analysis, a procedure for summarizing and integrating research findings, is often used to determine whether the evidences suggest that a particular treatment works and, hence, support a particular intervention. Oftentimes, the means of studies are used to calculate a weighted grand mean that can be used to derive an effect size. We agree that meta-analysis can help us to make decisions about treatment intervention; however, some fundamental challenges remain until the results of meta-analysis can be directly applied to individual people. First we will need to better understand the person-by-treatment interaction problem (Bouffard, 1993). Second, we will need to understand the problem of within-person variability, that is, why does the same person sometime react differently to the same treatment.

Generalizing from research results to a person, instead of a group, is a complex issue about which we have not found good answers. If adapted physical educators want to implement interventions based on the best evidence available (whatever that is), they need to acknowledge the possibility of intervention failure: people may not improve or may even get worst. We need to be prepared for these cases and prevent the possibility of harmful interventions. The best solution to this significant issue might be to constantly monitor the intervention to obtain feedback (Lambert, 2010).

**Question 5: Are Most Research Results Universal or Contextual?**

Evidence-based practitioners assume that research results are generalizable. When discussing research assumptions Slife (1998) noted that people espousing a modernist worldview typically endorse the idea that research results are generalizable whereas postmodernists typically argue for the contextuality of results. We think that, in general, results from the natural sciences are more generalizable than results obtained from the social sciences. As eloquently noted by the physicist Stephen Weinberg, every hydrogen atom in the universe has characteristics identical to any other hydrogen atom. The same is not true for each and every person or social system. It cannot be assumed that all research results, irrespective of the “quality of inquiry,” are generalizable and hence useful knowledge that can be applied. Many research results, particularly in the human sciences, may only have contextual applications (Polkinghorne, 2004).

According to contextualism, propositions can be corroborated, or not, given the context. It is the task of researchers to determine what works in which contexts. Effectiveness studies implicitly recognize the plausibility of contextualism. The main purpose of these studies is to determine if the results of efficiency studies can be generalized to different contexts. We think that contextualism in APA is a real possibility. It would be dangerous to behave as if every person was like any other person. Again, the monitoring of intervention will be essential to determine the effectiveness of the intervention and avoid possible harmful effects.
Question 6: Should All Decisions About What Works Be Evidence-Based?

In Canada, we have a health care system that typically provides free access to physicians and numerous medical services. However, some treatments are not supported because they are not “evidence-based.” One example is the use of botox for migraine headaches. Despite a number of supporting studies, experts concur that we do not have sufficient evidence that botox works and, hence, patients pay the physician for its administration. The treatment might actually work using the highest standards of EBM (i.e., RCT, meta-analysis) but these studies have not been completed yet.

We need to remember that typical evidence-based research can establish whether a treatment is better, on average, than another treatment or a placebo. However, a nontested treatment might still be superior to whatever has been supported via evidence-based research. The nontested treatment is simply “not empirically supported yet.” Consequently, a rigid adherence to the idea that we need to only provide evidence-based interventions may deprive users of optimal treatment. We submit that if there are good reasons to believe that a nonempirically supported intervention can produce beneficial effects, then the decision to use it or not must be contextual and based on assessment of the costs and benefits. Of course, if the intervention is implemented, careful monitoring of the situation must occur.

Question 7: Why Is Evidence-Based Practice not Evidence-Based?

Supporters of EBM have often been asked to provide the kind of evidences that they demand of medical practitioners and, surprisingly, at the time of this writing, there is little evidence that EBP works (Cohen, Stavri, & Hersh, 2004; Dobbie, Schneider, Anderson, & Littlefield, 2000; Haynes, 2002; Sehon & Stanley, 2003; Thorp, 2007). As recognized by Haynes (2004), one advocate of the EBM movement,

a fundamental assumption of EBM is that clinicians whose practices are based on an understanding of evidence from applied health care research will provide superior patient care, compared with that of others who rely on understanding of basic mechanisms and their own experience. So far, no convincing direct evidence exists to show that this assumption is correct. (p. 228)

Overall, EBP is not evidence-based and it does not meet its own tests of efficiency and effectiveness. We are aware of only one study that attempted to determine the effectiveness of EBP (Gülmezoglu et al., 2007), and the authors reported that “no substantive changes in clinical practices were detected” (p. 16).

In 1992, the EBMWG anticipated that the “proof [of superiority of EBM] is no more achievable for the new paradigm than it is for the old” (p. 2424). In an editorial published in the British Medical Journal, Reilly (2004) acknowledged the lack of evidence supporting the movement and further stated that EBM cannot be supported empirically “because, as a general proposition, it cannot be proved empirically” (p. 992). This lack of evidence did not stop him from stating that “anyone in medicine today who does not believe it [i.e., EBM] is in the wrong business” (Reilly, 2004, p.
In addition, others, such as Goodman (2003) and Thyer (2008) have claimed that not to practice EBP is unethical despite its unknown effectiveness. Briefly, the advanced argumentation has taken on the “tones of a moral imperative” (Buetow, Upshur, Miles, & Loughlin, 2006, p. 401).

The lack of evidence about EBP is one of the most disconcerting states of affairs about this movement. Contrary to the opinion expressed by the EBMWG (1992) and Reilly (2004), we think that state-of-the-art design and data analysis strategy can provide a partial, although never certain, answer to this crucial question. A thorough assessment of this question will, most likely, require a combination of qualitative and quantitative methods (Thorpe, 2007). Briefly, we contend that, if EBP is implemented into APA, we need to determine whether it works better than the best alternative approach.

**Question 8: Should We Endorse Current Governing Practices?**

In some parts of the world, the EBP movement has led to the development of associations with major organizations, including governments, and the creation of new organizations. These organizations have affected the development of EBP in education. In this section, we outline how some of these associations and organization have developed in the United Kingdom and the United States of America and question some of the observed trends. A good understanding of EBP is not possible without an appreciation of the power of these organizations. They have affected “what counts for legitimate knowledge, what research is funded, the standards for training of young scholars, and the studies taken into account in policy formation” (Walters & Lareau, 2009, pp. 2–3).

Although medicine and education are different fields, the development of EBP in education over the last 15 years has mirrored key ideas advanced in EBM. Suggestions that educational research should emulate medicine were made by Slavin in his Distinguished Lecture presented at the American Educational Research Association annual meeting in 2002. He contended that the “most important reason for the extraordinary advances in medicine, agriculture, and other fields is the acceptance by practitioners of evidence as the basis for practice. In particular, it is the randomized clinical trial—more than any simple medical breakthrough—that has transformed medicine” (Slavin, 2002, p. 16).

During the late 1990s, it was frequently stated that education research was of low quality and relevance to policy makers and practitioners (Hammersley, 2007; Walters, Lareau, & Ranis, 2009). Hence, both its quality and utility needed to be improved (Ranis, 2009). As a reaction to these perceived deficiencies, new organizations, policies, and even laws were developed in some countries. As in EBM, some advocates of EBP in education have proposed hierarchies of knowledge stipulating that RCT and meta-analysis are the tools of choice to answer fundamental questions in education (Hammersley, 2007; Walters, Lareau, & Ranis, 2009).

We find the developments observed in the United Kingdom and the United States over the last 15 years to be particularly instructive (for elaboration see Hammersley, 2007; Walters, Lareau, & Ranis, 2009), and we review them to illustrate some alliances that have been formed between advocates of EBP, governments, and major organizations.
In the United Kingdom, the New Labor Party came into power in 1997 and declared that a more effective use of research could be made to address policy issues. In the 1997 party’s manifesto, it could be read that “What counts is what works” (Labour Party, 1997), and what works can be determined by conducting research. In numerous domains, including education, the manifesto said that research should be used to guide policy development and practice. Criticisms of the educational system were also made from the inside. Examples of these criticisms are provided in Hargreaves (1996), the Tooley Report (Tooley & Darby, 1998), and the Hillage Report (Hillage, Pearson, Anderson, & Tamkin, 1998). Hargreaves (1996), who was particularly incisive, advanced that most education research was of poor quality and of limited relevance and that it needed a “very different kind of research if it is to influence practice” (1996 [2007], p. 5). Drawing a parallel between medicine and education, he claimed that much medical research is not itself basic research “but a type of applied research which gathers evidence about what works in what circumstances” (1996, p. 5) and argued that educational inquiry and practice should follow this model. Hargreaves suggested that education should model itself on recently developed ideas in EBM in which clinical decision making is based and justified in terms of research findings about the relative efficiency and effectiveness of different medical treatments. To him, education needs a better integration of research and practice and more research is needed about what works, for whom, and under what circumstances.

Hargreaves’ allegations about the low quality of educational research were subsequently supported by government sponsored reports, including one to the Economic and Social Research Council’s Teaching and Learning program (McIntyre & McIntyre, 1999) and another by the Institute of Employment Studies (Hillage, Pearson, Anderson, & Tamkin, 1998). Briefly, internal criticisms of educational research and the political climate at the time fostered the development of EBP in education.

A similar academic debate began during the mid 1990s in the United States where a number of reports claimed that most educational research was unsound because it was based on inferior research designs and shoddy data. Hence, the results were not trustworthy (Walters & Lareau, 2009). At the same time, it was argued that educational policy and interventions ought to be based on sound research (Walters & Lareau, 2009). The United States Congress responded to these allegations by adopting a number of laws (e.g., No Child Left Behind Act [NCLBA], Education Science Reform Act of 2002). However, critiques of educational research and the Congress advanced a particular view of quality research. Of particular importance to the present discussion are the definitions of “scientifically based research” adopted in NCLBA (see Appendix) and the Education Science Reform Act of 2002. These laws stipulated what scientifically-based research in education is and what is meant by scientific validity. It was advanced that the “gold standard” of educational research is the RCT.

The NCLBA created an impetus for basing educational practice on scientifically based research. On November 2002, the Institute of Education Science (IES), a new division of the United States Department of Education (USDE), was established by President George W. Bush with Grover J. Whitehurst as the director of IES for a period of six years. The IES put into place a structure that defined, funded, and disseminated what counts as quality educational research and received $18.5
million to develop the What Works Clearinghouse (WWC). On the WWC site, it is asserted that the Clearinghouse “is a central and trusted source of scientific evidence for what works in education,” which “assess the rigor of research evidence on the effectiveness of interventions (programs, products, practices, and policies)” and “develops and implements standards for reviewing and synthesizing educational research” (What Works Clearinghouse, n.d.). Following recommendations made by the Coalition for Evidence-Based Policy, the USDE has established that RCT will receive priority funding in the evaluation of educationally-based interventions, as opposed to research involving other designs (Thyer, 2008). On the WWC site, one may read that currently, “only well designed and well implemented randomized controlled trials (RCTs) are considered strong evidence” (p. 12). A major goal of NCLBA has been to guide policy makers and service deliverers in their use of interventions. To receive funding, priority is given to schools that implement intervention strategies supported by scientifically-based research.

The federal policies and guidelines about what constitute good educational research methods were supported by considerable federal funding. For example, it is stated in the Department of Education’s 2002–2007 strategic plan that one of its goals was that 75% of its funded research addressing causal claims should use random assignment by 2004 (U. S. Department of Education, 2002). In addition, during the fiscal year 2002, approximately 74% of all IES projects used a random assignment of participants to groups and all these projects addressed causal claims (U. S. Department of Education, 2003). In summary, governments and other major institutions can determine what can be studied and discussed, how research ought to be done (Cat, 2006) and, more importantly, it can legitimize what is accepted as true (Foucault, 2007).

The policies of the USDE and the IES have been criticized on a number of grounds. First, only topics amenable to true experiments (Shadish et al., 2002) receive priority funding. Second, contrary to an assumption made by most researchers that methods are tools to answer research questions, methods drive the inquiry process. Third, the policies limit inquiry to the logic of justification and give a limited, or nonexistent, role to the logic of discovery or creativity. More fundamentally, fourth, the IES apparently assumes that key ontological or epistemological assumptions have either been settled or are unworthy of discussion. Why deal with these “inherently unresolvable philosophical issues” (Thyer, 2008, p. 342)?

We are not aware of other legislation stipulating how educational inquiry ought to be conducted around the world. Some reevaluation of educational research, however, might be underway outside of the USA and the United Kingdom given worldwide movements toward accountability and the numerous calls for EBP (Wiseman, 2010). We agree that governments, researchers, policy makers, and practitioners should engage into discussion or debates about research priorities; however, we contend that decisions about how research ought to be conducted are better left to researchers interacting with participants.

**Question 9: Shall the Domain of Acceptable Inquiry Be Restricted?**

The NCLBA and the development of organizations like the Cochrane collaboration, the What Works Clearinghouse, the policies of the IES, and the EBP movement...
in general have prioritized the use of the RCT. As noted earlier, different hierarchies of evidence explicitly state that qualitative inquiry is of low value to answer evidence-based type of questions. Although these methods are typically viewed as useful, they do not achieve the status of the gold standard of inquiry, the RCT.

Needless to say, this trend has generated strong opposition and has reopened the qualitative-quantitative debate (Lather, 2004; Morse, 2006a, 2006b; St. Pierre, 2004, 2006). These concerns were well summarized by Morse (2006a):

The new emphasis on evidence-based care has undoubtedly repositioned qualitative research as second class science—again. When funding agencies prioritize research according to the Cochrane (1972/1989) criteria, qualitative research is not fundable; when medicine defines rigor by these standards, we take the last position; our methods and approach are classified as mere opinion. But the sad consequence of this perspective is that the method perceived as rigorous and fundable then shapes the substantive research agenda. Researchers who are used to seeing the world in terms of clinical drug trials or epidemiological designs do not consider the type of research questions that qualitative researchers are interested in as “researchable,” significant, relevant, or even on their radar screen. My point is that, unfortunately, research gets on a trajectory that includes a specific type of question that the current methodology in vogue best answers and other problems are left on the back burner or ignored. (p. 315)

To some this debate shows that there “is an ongoing controversy over the political economy of evidence, how it is defined, and who defines it. This is not a question of evidence or no evidence but who controls the definition of evidence and which kind is acceptable to whom” (Larner, 2004, p. 20). In summary, the “stakes in this renewed, ongoing, and rapidly escalating debate are high—intellectually, socially, and educationally” (Phillips, 2006, pp. 3–4).

Adapted physical activity is a multidisciplinary field and we believe that there are numerous possibilities for interdisciplinary inquiry. Researchers conducting multidisciplinary research conduct their own disciplinary “things”; each member of the research team acts as a disciplinary expert. Knowledge claims are based on the authority of that expertise, and the final product does not integrate the separate disciplinary contributions. In interdisciplinary research, researchers may choose research problems, which require extensive insights from several disciplines; they must have a full and colloquial understanding of the major portion of the second discipline. In interdisciplinary research, an attempt to integrate the disciplinary contributions is made (Miller et al., 2008). If research methods drive the inquiry process, then the range of significant disciplinary, multidisciplinary, and interdisciplinary contributions is reduced. Only research questions amenable to methods at the top of the hierarchy are legitimate and, hence, other perspectives about research questions might be rejected. This could have significant consequences for the future of our field including more fragmentation.

**Question 10: Will Evidence-Based Practice Work?**

A fundamental assumption of advocates of EBP is that it will lead to better policies and practices. For a number of reasons, we must remain skeptical about this assertion. First, it has not received empirical support. In fact, proponents of the
EBMWG (1992, p. 2424) declared it untestable (see Question 7). Second, the results obtained using RCT and meta-analysis may have limited generalizability. For example, LeLorier, Grégoire, Benhaddad, Lapierre, and Derderian (1997) compared the results of 12 large RCT with the results of meta-analyses published earlier on the same topics in four prestigious medical journals. The outcomes of the 12 RCT were not predicted accurately 35% of the time by the meta-analyses. This replicability issue suggests that we must be cautious when using the findings of RCT or meta-analysis. If the results have limited replicability when obtained by these two highly ranked methods of EBM, one can theoretically expect that the generalizability of findings to be lower when confronted with nonlaboratory situations where numerous variables potentially affecting outcomes are not under control. Third, we must remember that randomization is a process that may not work (Worrall, 2010). An interesting example was produced by Leibovici (2001) who identified 3,393 participants with a bloodstream infection. The participants were randomly assigned to a treatment group and a control group. At the beginning of the study, a check was made for baseline imbalance related to risk factors and severity of illness. None was found and, hence, it was claimed that the groups were equivalent at the beginning of the study. The names of those in the treatment group were given to a person “who said a short prayer for the well-being and full recovery of the group as a whole” (p. 1450). The mortality rates were not statistically significant. However, both length of stay in hospital and duration of fever were shorter in the intervention group. In accordance with the logic of RCT, Leibovici recommended that intercessory prayer “should be considered for clinical practice” (p. 1451). Leibovici’s trial was a cautionary demonstration that a large RCT could support a theoretically implausible intervention (Leibovici, 2002). Fourth, adapted physical activity is a field guided by a philosophy of individuality. That is, most of our offerings focus on perceived needs and interests of each person. In this context, it is prudent to remember that the typical statistical analysis of data in a RCT allow the inference from sample aggregate to population aggregate; however, the inference from population aggregate to a person is not justified (Bouffard, 1993). More fundamentally, fifth, EBP assume that, to a certain extent, the future is predictable and high quality research can help us predict what is likely to happen when an intervention is made. We submit that this assumption is questionable (for elaboration see Rescher, 1998). Our skepticism is based on the following general systems argument. All living systems are open systems because they must exchange matter and energy with their environment to survive. In all open systems the outcome of a process cannot be predicted from the initial state of the system (von Bertalanffy, 1950). Although the systems studied in the biological sciences typically respond to matter, energy, and information, these systems are, to the best of our knowledge, largely unaffected by meaning and interpretation. In the psycho-social sciences, participants can often alter their response based on the meaning or interpretation they give to a situation. Hence, the possibility of predicting what works in the psycho-social sciences appears even lower than in the biological science. Of course, the presented argument does not imply that prediction in the psycho-social sciences is impossible. It is, however, congruent with numerous studies reporting a low proportion of variance predicted.12

The argument presented in relation to Question 10 does not mean that evidence should not be used. It means that we should be cautious when using them.
Even carefully planned interventions may not produce the desired effect and lead to negative outcomes. Even if we use the best evidence available (whatever that is), we will need to regularly monitor interventions and be ready to alter or stop them (Lambert, 2010).

**Summary and Conclusion**

We have argued that the many definitions of EBP reflect the particularities and the histories of different professions. It is not possible to provide a succinct “one way fits all” definition of this movement. At a most general, abstract level, we can find some commonality among the different uses of the term (e.g., it refers to the use of best evidence to support practices; it is a technological viewpoint). However, as soon as more complex questions are asked (e.g., What is evidence? How do we rank evidence? How do we integrate evidence? What is a practice?), different answers are provided. Although some may dislike this elusive concept, we think it is a plus. Instead of borrowing a rigid conceptualization of EBP to develop practices that may be incompatible with some of the major goals of APA, we have the opportunity to develop our own EBP without being constrained by a rigid viewpoint.

Although we have observed some enthusiasm about EBP in publications (Hutzler, 2006, 2010, 2011; Jin & Yun, 2010), and during informal discussions with our colleagues, we contend that a cautious attitude toward EBP is needed to avoid some of its potential pitfalls. The movement raises a number of complex ontological and epistemological questions that are often ignored. If we are not cautious, our field might inherit some problems from the past. The “originators of EBM paid little attention to the philosophy of science, and attempts to do so now are mainly post hoc and perhaps defensive for some EBM advocates” (Haynes, 2004, p. 235). We contend that little attention has been given to fundamental research assumptions (Slife, 1998), and engagement with these issues must be done to develop our own EBP. Although it is often claimed that scientists are not inclined to engage into philosophical issues (no evidence!), science and EBP cannot be divorced from philosophy. In fact, traditional philosophical questions are embedded into the EBP debate.

Hierarchies of evidence are at the center of a number of questions we raised. They are “applied epistemologies” and like any epistemology they need to be justified. The development or hierarchies of evidence can easily reopen the qualitative-quantitative debate and the “science war” within APA. Tacit, or explicit, worldviews are embedded within hierarchies that ultimately promote some research methods and reject others. History tells us that, in most cases, hierarchies are the results of expert judgments; however, the substantive rationale behind these hierarchies is not provided and, hence, the evaluation of ideas is impaired. If we develop hierarchies of knowledge production in APA, we also need to clearly articulate the rationale behind them. To state that a consensus of experts was achieved will be insufficient. We think that epistemological pluralism will be essential to avoid opening acrimonious and divisive debates within our own field. Epistemological pluralists prefer the use of multiple approaches to knowing and are distrustful of over-reliance on any one epistemological approach. They believe that multiple approaches to knowing are needed to better approximate truth (Marshall, 1998).
We think that the “model” presented in the Winter 2005 issue of *Exceptional Children* is leaning toward epistemological pluralism and it deserves careful scrutiny in our field. In this issue, the authors attempted to portray a broad view of EBP and research in special education. The authors acknowledge the complexity of special education research and outlined guidelines for four types of research: experimental and quasi-experimental (Gersten et al., 2005), single-subject (Horner et al., 2005), correlational (Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005), and qualitative (Brantlinger, Jimenez, Klingner, Pugach, & Richardson, 2005). They proposed standards for EBP using experimental and quasi-experimental research (Gersten et al., 2005) and single subject research (Horner et al., 2005); however, equivalent standards were not developed for the two other types of inquiry. We applaud the acknowledgment that different methods are needed to address different questions. As noted by Feuer, Towne, and Shavelson (2002), “if a research conjecture or hypothesis can withstand scrutiny by multiple methods, its credibility is enhanced greatly” (p. 8). It is important to note that the 2005 guidelines and standards have been developed for some research methods. The more challenging task of developing guidelines and standards taking into account different methods remains to be completed.

We have argued that EBP and politics have converged and created a dangerous mix. Evidence-based practice ideas have been inserted in the NCLBA act to define what scientifically based research is. We submit that our field should adhere to a broader view of inquiry. The EBP movement is fundamentally guided by one question: What works? We accept the relevance of this question to answer pressing needs of our society; however, the focus on what works can easily overshadow the “why?” and “how?” questions. This focus should not lead to an abandonment of these traditional and fundamental research questions. Although we recognize the role that ought to be played by governments and major research organizations in establishing research priorities, we submit that the decisions about how to conduct inquiry are better left to the researchers.

In this paper, we introduced the reader to 10 questions about EBP. Our coverage of fundamental questions is incomplete. We barely discussed knowledge integration issues, and we did not address questions related to knowledge translation. In particular, we did not discuss challenges related to the use of evidence by the service provider and, as well, the preferences of the participant that we submit should be empowered and self-determined (for elaboration see Spring, 2007). In addition, our ability to predict participants’ responses to interventions is still fairly limited, and we submit that whatever EBP the field develops, we will need to constantly monitor our interventions (Lambert, 2010). More fundamentally, the development of EBP can challenge some conceptualizations and understandings of APA (Hutzler & Sherrill, 2007; Reid, 2003). Many EBP ideas have their roots in medicine—this is why we borrowed extensively from this field to write this paper. Recently, the dominance of the medical understanding of disability has been questioned by those who champion a sociocultural understanding (e.g., DePauw, 2000; Grenier, 2007). Typically, proponents of medical models and social models of disability make different assumptions about the causes of disability (see Smith, 2009, Table 1). Proponents of medical models typically argue that disability is “caused by fixed medical characteristic” (Smith, 2009, p. 22), while proponents of social models argue that the cause of disability is social and, hence, interventions
should change the social and political environment (Smith, 2009). We think that EBP, as currently conceptualized in a number of fields, has more affinities with medical models of disability. It is unclear to us how social models and EBP are compatible. Finally, we reemphasize that EBP is fundamentally based on the idea of conducting technological interventions (Standal, 2008). Interventions are made to reach goals—they are means to reach ends. Oftentimes, physical activity is practiced “for its own sake” instead of a tool to reach an end. In this case, the goals of APA and EBP may be incompatible (for an enlightening discussion of this issue see Standal, 2008).

Given the strength of the EBP movement, and our informal observations at adapted physical activity conferences around the world, we anticipate that the ideas discussed in this paper will soon, if not already, have a major impact on the field. We hope that our presentation has convinced you that this apparently simple idea has had numerous ramifications and, hence, we need to carefully step forward and anticipate consequences of action. As noted by a colleague after a presentation of an earlier version of this paper at a research conference, “there is an elephant in the room” and we should not ignore it. We need to prepare our field to respond to this important trend.

Although we have raised a number of red flags about EBP, we have the benefits of hindsight. Major developments in EBP have mainly occurred during the last 20 years, and the 10 questions we raised reflect some of the issues we have observed about this important movement. We believe that we can develop our own EBP such that we alleviate some errors of the past. To avoid these pitfalls, we need to be cognizant of them. Ultimately, we hope that this paper has been a consciousness-raiser (Slife, 1998). An awareness of problems and issues is the first step to solve them.

We know that some of you will not be comfortable with all our questions. To prepare the future, let us engage in a professional, unfettered, and constructive debate. The stakes are high!

End Notes

1Terry Rizzo was the Editor for this manuscript.

2This site is no longer available. The definitions can be obtained by contacting the first author.

3In evidence-based healthcare, the focus is often the population as opposed to the person. In this context, the broad definition outlined does not apply.

4By focusing on what works, fundamental why or how questions have been overshadowed by the EBP movement. We may know what works and, at the same time, have little knowledge about why or how it works. The focus on what works means that RCT have mainly been used to produce causal description as opposed to causal explanation. However, in a RCT it is possible to include the search for moderators and mediators (Kraemer, Wilson, Fairburn, & Agras, 2002) as well as mechanisms of change (Kazdin, 2007).

5In preparation for this paper, we consulted numerous hierarchies of evidence sites and never found a comprehensive rationale for hierarchies. If any reader finds a comprehensive rationale justifying a hierarchy, please contact us. We will gladly withdraw this statement. Our key concern here is that without a rationale, there is no room for debate and discussion.

6We do not claim that similar associations and organizations exist in other countries. Our purpose is to illustrate ideas advanced by these organizations and how they have affected EBP in these countries.
Although the manifesto refers to research, we submit that the Labor party in fact argued for the more frequent use of program evaluation. For a distinction between research and evaluation see Smith and Glass (1987).

On this site, we can also read that “quasi-experimental designs . . . with equating may only meet standards with reservation; evidence standards for regression discontinuity and single-case designs are under development” (What Works Clearinghouse, 2008).

In the 2007–2012 strategic plan, one of the reported objectives of the U.S. Department of education is to “[t]ransform education into an evidence-based field” (U.S. Department of Education, 2007, p. 17).

A variety of expressions like multidisciplinarity, interdisciplinarity, transdisciplinarity, and crossdisciplinarity have been advanced to refer to different types of inquiry involving two or more disciplines (Miller et al., 2008; Sherrill, 2004; Strober, 2011). Unfortunately, these words are often used interchangeably or have different meanings. It is beyond the scope of this paper to outline the different definitions. Our usage of multidisciplinarity and interdisciplinarity is presented in this section. We think it is sufficient for the purpose of this manuscript.

Believers think that there is a good theory guiding the study and, hence, argue that this study, and other similar studies, provide evidence that intercessory prayer works. For an interesting and illuminating discussion of intercessory prayer see Giacomini (2009).

Multiple regression analysis is often used to predict the proportion of variance accounted in a criterion variable. We need to remember that the weights selected via a regression procedure maximize the correlation between the criterion and a linear combination of predictors. When the same weights are applied to another sample, the accuracy of prediction is usually lower.

The science war was essentially a series of intellectual debates during the 1990s between scientific realists and postmodernists about the nature of scientific theories. During this decade, a particularly acrimonious debate followed the publication of two manuscripts by the physicist Alan Sokal. His first paper entitled “Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity” was published in Social Text. When this paper was accepted for publication, he published another in Lingua Franca revealing that the first paper was a hoax.

The proposed criteria and standards were recently evaluated by Cook, Tankersley, and Landrum (2009).

References


Haynes, R.B. (2002). What kind of evidence is it that Evidence-Based Medicine advocates want health care providers and consumers to pay attention to? *BMC Health Services Research*, 2, 3.


**Appendix**

**Definition of Scientifically Based Research in No Child Left Behind Act**

The term “scientifically based research”

(A) means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs; and

(B) includes research that:

(i) employs systematic, empirical methods that draw on observation or experiment;

(ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;

(iii) relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;

(iv) is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls;

(v) ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their findings; and
(vi) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.