When Best Practices Sometimes Aren't

Peter Hlebowitsh
College of Education
The University of Iowa

The Iowa Academy of Education Occasional
Research Paper #9
Added: July 2011

Much of the recent talk about school reform has focused on the difference-making power of the teacher in the school lives of children. Because teachers make an obviously significant contribution to student learning, we are now experiencing an increasing research effort to identify the practices that successful teachers use in the interests of making them available as exemplars to other teachers. In this way, the teachers who aspire to succeed in the classroom can follow in the wake of the work performed by their presumably more enlightened counterparts. The idea here is reasonable: Successful teachers must be doing things that all teachers should consider doing, so let's identify their practices, distill them into a framework that we might call "best practices" and find a way to bring these insights to the over three million teachers who populate our public schools.

The problem, of course, is that certifying a construct that we comfortably point to as "best practices" is not as easy as it might sound. It requires establishing genuine causal or functional linkages between a teacher's instructional behavior and measured achievement. And the best way to find such causal effects is through an experimental research framework. As a consequence, 'best practices' are often drawn from studies that have used an experimental research design, which explains why they are also commonly known as evidence-based practices, or even in some cases, as scientifically-

based practices, or even what the federal government sometimes calls *what works*.

Evidence-based practices certainly have their appeal, especially as one considers them against the ideologically-driven thinking that marked many past efforts to bring reform and change to the schools. Dualistic ideological debates, for instance, over whether phonic or whole language approaches are best for young children are thankfully less likely to prevail in a context that demands evidence. To its advocates, evidence-based practice represents an advancement toward the creation of an inventory of instructional strategies, cataloged along key analytical lines, that might allow educators to ask whether practice x or program y has any record in paying a general achievement dividend in any number of key academic core areas (reading representing the most important one), or with particular subgroups of children. Or even if practice x or program y might represent some way to better advance other key learning outcomes, both cognitive and non-cognitive.

This essay, however, raises questions over whether the logic of best methods is even appropriate to the educational situation. Because teachers draw from multiples sources of evidence to deal with a variety of different classroom-based problems, the notion of identifying a general set of best

methods that holds for particularized teaching problems and teaching contexts is dubious (Spillane and Meile, 2007). In the research community, such a problem is known as the ecological fallacy, which is an idea that cautions us against concluding that reliable inferences can be made about specific individuals based on aggregate data collected from the group to which those individuals belong. Even if pitched at a level of high generality, evidence-based insights about teaching still have the difficult task of finding particular relevance in the dynamic contextual complexities of the educational situation. The research studies used to generate evidence-based practices, we should remember, speak from the perspective of central tendencies. The act of teaching, however, always inherits a local condition, with variations that cut across and within schools and classrooms, making it difficult for educators to align their own particular instructional dynamic with what others are saying is best form an evidence-based perspective. Clearly, there is error (or fallacy) in assuming that individual members of a group necessarily carry the average characteristics of the aggregate group. Most distressing is the prospect of assuming that best teaching practices are reliable and portable enough to script instruction, which poses the harmful possibility of closing down the discretionary space teachers need to make responsive and educationally sound judgments in the classroom.

Teaching to a Central Tendency

The No Child Left Behind Act (2001) defines programs validated through scientifically-based research as needing to fulfill the requirements of scientific rigor. This means that the studies used to support educational interventions must be experimental, replicable, and based on the collection of reliable and valid data. Ideally, the programs and practices supported by scientifically-based research are less committed to identifying a federally "approved list" of reading programs that schools could use and more interested in allowing states and districts to make choices about instruction, as long as the programs and materials pass the test of scientific rigor. The thinking here is that as teachers and principals contemplate ways to go about providing interventions that will address various instructional problems, they might use the principles of scientifically-based research to find some answers. Thus, if enough studies support one-on-one tutoring for at-risk readers in grades 1-3 by showing that the average tutored student reads more proficiently than the untutored students in the control group, one might determine that such a program might be a good fit for one's own school or classroom. And in this sense, these are the kind of data that can indeed inform a teacher's instructional

judgment or a principal's desire to develop or enhance certain school wide outcomes.

But such studies (and the general faith put into this rationale for school improvement) have a long way to go before we can be certain that they can yield best method approaches or even evidence-based approaches in the classroom. The problem fundamentally has to do with having unreasonable faith in the dependent variable used to scientifically demonstrate desirable interventions, or as Spillane and Meile (2007, p.48) noted, faith in the tendency to convert a single piece of information (observing that, say, test scores, for the experimental group are, on the average, significantly better than those achieved by the control group) into a single methodological conclusion (deciding that the instruction featured in the experimental group must be good for all).

Teachers, especially teachers of readings, might find some advantage in knowing that certain programs and methods pay some general achievement effect, but the viability and vitality of such programs or practices will ultimately be tested in the heat of the classroom – in the particularized dimensions of the teaching situation. Thus, instructional practice is, in a manner of speaking, resistant to advice drawn from group or aggregate effects. Teachers should not be in the spot of paying more attention to what the research tells them over what their personal experiences and situational judgment tell them. Research looks for

the line-of-best-fit on the scatterplot, whereas the teacher is looking for the way or ways to teach all the children represented on the scatterplot.

This is another way of saying that the idea of evidence is situationally constructed by teachers and is not reducible to a single dependent variable taken from a single shot measurement. A Spillane and Miele (2007) observe, "school leaders and teachers use different sorts of information to construct evidence, including published research reports, word of mouth accounts of what works, student achievement scores, demographic data and personal experience" (p.48). The implication is that operational answers to good teaching cannot be drawn from research studies that identify practices or methodologies believed to be portable to all classrooms. The answers are in the emergent judgments of the teacher, who is naturally obligated to be apprised of the research but who also understands that the "right" decision in a classroom depends on weighing particularistic factors related to the nature of the child, to available resources, to the defined purposes in the curriculum, to the subject matter at hand and to a raft of other variables residing in the educational situation. The eminent education scholar, Joseph Schwab, described the classroom situation.

> There are thousands of ingenious ways in which commands on what and how to teach can, will and must be modified or circumvented in the actual moments of teaching… Moments of choice of what to do, how to do it, and whom and at what pace, arise hundreds of times a school day and arise differently every

day and with every group of students. No command or instruction can be so formulated as to control that kind of artistic judgment and behavior, with its demand for frequent, instant choices of ways to meet an ever varying situation. (p. 245)

Others, of course, are also looking for this more complicated construction of good teaching. For instance, Howard Gardner, whose work has helped to expand our ideas on how educators construct intelligence in the classroom, is much less inclined to talk about evidence-based teaching than about personalized teaching; that is, teaching that accounts for different learning strengths and that recognizes different profiles of individual potential and capacity. According to Gardener, good teachers make differential use of multiple methodologies, multiple assignments, multiple resources and are on the prowl for alternative pedagogical possibilities as they might emerge from the educational situation. "Anything worth teaching," observed Gardner, "usually can be taught in a number of ways; by using multiple forms of representation and presentation you reach far more students" (Rules for…2007).

Even the medical community, which is often used in analogy to support evidence-based judgment, is not is full accord on the matter of evidence-based thinking. The medical professions are keen to represent the judgment of their practitioners as based on evidence drawn from only the highest experimental

standards of research. But critics have asserted that such a principle has come at the expense of practical judgment in the clinical setting. The problem, as Tonelli (2001) argues it, is that experimental evidence taken from clinical trials does not always directly apply to individual patients. Tonelli encapsulates the point by noting that evidence-based practice in medicine has produced "clinical practice that has subtly shifted away from the care of individuals toward the care of populations" (quoted in Brusling, 2005, p. 93) To be sure, medical practitioners should be informed by experimental evidence, but other sources of insight not accounted for in the evidence-based rationale (including physiologic principles, patient and professional values, and system features) also come into play. Adding to the complexity is the reality that each of these sources have weights that vary from case to case.

Obviously, good teachers, (like good medical practitioners) do "more than follow authoritative evidence-based rules for practice" (Brusling, 2005). Acknowledging this truth requires more humility over what evidence-based instruction can deliver, including less assertive references to "what works" and some concession on moving the discourse about teaching away from a binary procedure of evidence-based versus not evidence-based, to something more like evidence-informed teaching or the more general question of whether teaching can even be treated as evidence-based (Brusling, 2005).

Evidence-based teaching is, by its very nature, largely a convergent exercise – a delimiting of different approaches in the interests of finding the golden mean. It opts for ideas that are centripetal in nature, rather than centrifugal, seeking to find an all-embracing and centered way to proceed. But even if we had solid data on the central tendencies found with the use of particular methods or programs on various cognitive goals (on, say, reading comprehension), we still could not know if they will work for any particular student or if they might otherwise come at the expense of other (often non-measured) factors, such as, for example, attitude toward learning or attitude toward reading itself.

The Historical Misconstruction of Effective Teaching

Evidence-based instruction historically derives from a process-product model of research aimed at finding some relationship between what teachers do in the classroom (process) and how well students learn in the classroom (product). The intention of the process/product design is to find some general linkage between discrete teacher behaviors and student achievement, all in the interests of identifying generically *effective* teaching strategies for all teachers to follow and to be supervised against. The signal hope was to find a constellation of teacher behaviors that possessed good correlates with

10

achievement, resulting in the identification of generically effective characteristics in the classroom, such as, say, teacher clarity, variability in the use of materials and techniques, type of feedback given and so forth. Thus, the term effective has been and continued to be largely defined as *conduct that lifts test scores.*

The search for the holy grail of generically effective teaching strategies started in earnest during the early parts of the late 1960's and early 1970s. Because the process product design was usually correlational, difficulties ensued over whether a characteristic such as, say, pupil involvement in a lesson, was actually part of some broader experience we can call effective or whether it had independent status as a characteristic of effective teaching. Nevertheless, researchers went forward with the project of stipulating the classroom (instructional) behaviors that were in large frequency in classrooms where the achievement levels were high.

When research is focused at finding process factors in relation to achievement, the derived insight drawn from the work tends to fix on the characteristics of particularized ways teaching. The result is that a teacher's status as an effective teacher is associated with an ability to activate general traits known to have some connection to measured achievement.

The nature of the so called teacher effectiveness literature, which was quite popular in the 1980's, illustrates the important point. Process/product researchers, for instance, have historically touted the importance of *time on task*, also known as *engaged time,* in the classroom. The effectiveness research found that instructional judgments that kept students on task (or engaged) were judgments sanctioned as effective because engagement itself was positively related to achievement. Such a finding is obviously unremarkable, in that it espouses the need for teachers to keep learners engaged in classroom activities, which most educators understand implicitly. But when such a principle stands at the forefront of how one views good teaching, the effects in the classroom can be less than enlightened. Because all forms of engagement are not necessarily educative, the *time on task* dictum leaves us wondering about the qualitative character of the task. And as such, it is of limited value. Thus, a teacher might achieve high grades from supervisors in keeping children on task, but if the task itself is not educationally worthwhile, or if the task has no good linkage to key purposes in the school curriculum, the issue of student engagement is moot. The issue of teaching cannot be easily reduced to only instructional manipulations.

Another manifestation of the same problem occurs when research attempts to identify and promote the value of certain universally 'effective' (qua, 'best') instructional practices. Research findings from the process/product

approach tend to indicate (to name a few generalizations) that effective teachers have high expectations for performance, that they convey enthusiasm in their teaching, and that they are vigilant about monitoring student work. It all sounds reasonable. But when the term 'effective' is used in this context, one needs to be able to show the criteria by which such a term is being operationally defined. And this is where the limitations to identifying what effective teachers do are most obvious.

We should all be reminded of the fact that the term 'effective' (or even 'best') has no intrinsic moral valence. Thus, one could be 'effective' (or best) at any number of things, including, say, thievery or murder. Similarly, the term effective can also be impartial to a good theory on what it means to be a professional educator. Researchers could, in effect, define effective as they please. And if the definition turns in the direction of raising standardized test scores, which has been the tradition, one can begin to appreciate the problems that follow because it will not take long to realize that what one might do to raise test scores may not always lead to enlightened teaching. For instance, in 1979, one teacher effectiveness researcher, whose work was dedicated to finding characteristic behaviors among effective teachers of disadvantaged pupils in the primary grades, concluded that

Teachers who ask the most high level and the fewest low level questions, teachers whose pupils ask the more questions and get more feedback from their teachers, teachers who tend to amplify or discuss pupil initiated comments most *are the ones who are least effective*. Teachers who use more low level questions and fewer high level ones, whose pupils initiate fewer questions and get less feedback, who tend not to amplify or discuss what pupils say- *these are the most effective ones* (Medley, 1979, p.24).

Similarly, another teacher effectiveness researcher, touting the achievement gains resulting from direct instruction for elementary school-aged children, concluded that "the frequency of nonacademic activities such as arts and crafts, reading stories to a group, or questions to students about their personal experience are usually negatively related to achievement gains" (Rosenshine, 1979, p.52).

The question to ask here is whether we can trust studies crafted against narrowly defined dependent variables enough to give us an appraisal of the classroom situation that results in some reliable declaration over what works or what constitutes a best method. The early effectiveness researchers, included the ones cited above, knew that their data could not take them to such heights, as they argued that much of the variance in teacher behavior was still not accounted for in their studies. They were cautious about how far one could take their finding into the decision-making realm of the classroom. The historical lesson is that the way that researchers operationally defined effective teaching may give sanction to behaviors that are at a considerable distance from

a sound theoretical profile of professional behavior. The current fashion for evidence-based teaching is no exception to the historical rule.

In other words, the assumption that standardized test measures represent a reliable and unimpeachable gauge for good teaching is clearly problematic. Most teachers understand and appreciate the value of standardized test scores, but their construction of what it means to teach well puts less significance on the scores than on student attitudes, values and behaviors in both academic and nonacademic endeavors. If teachers were asked to try to define evidence-based teaching, the dependent variable would necessarily be much more complex than any single-shot test score. The point is that the standards that teachers have for judging success in the classroom are broader and more intricate than anything emerging from the evidence-based research paradigm. If evidence-based teaching becomes the main criterion for classroom (instructional) decision-making, we could very well find that teachers will follow this less complex and less responsive path into the classroom.

Discretionary Space

The most detrimental effect of the failure to account for the complexity of the classroom is the loss of professional discretionary space for teachers, who might find that they have less opportunity to exercise their own intelligence and creativity in the classroom. Schools have a significant history

with this problem. During the competency-based movement of the 1970's, for instance, a popular method of curriculum development proudly yielded "teacher-proof" curriculum materials. The intention was to the design the curriculum in a way that reduced the teacher to the role of a functionary in the school, carrying out the specifications of the curriculum according to the orders laid down by curriculum planners. Curriculum objectives, lesson designs, practice activities and even student tests were prefashioned, given to the teacher to follow uncritically. In some cases, actual language was scripted for teachers to utter during their lessons with students.

The idea of scripting teacher behavior in the classroom is, to use Eisner's (1998) characterization, one that rests on the erroneous assumption that teachers are best viewed as subservient to the curriculum and that learning itself follows some linear assembly-line process. We know that attempts to apply formulaic approaches to teaching contradict the complex nature of the classroom and result in reducing the teacher's role to its most routine and rudimentary (and unprofessional) elements. As Dewey (1904) observed, "teachers should be given to understand that they are not only are permitted to act on their own initiative, but that they are expected to do so, and that their ability to take on a situation for themselves would be more important in judging them than their following any particular set method or scheme" (p.27-28).

So, how can we begin to frame effectiveness in relation to the issue of discretionary judgment? Most people, if asked to define effective teaching, would probably acknowledge the difficulty in answering the question while noting that they would know good teaching when they saw it.

And if they saw it, the portrayal will likely have two main features. First, teacher effectiveness would likely speak to outcome expectations in the classroom, meaning that teachers must produce results indicating that students have learned (among many other things) how to read, how to do mathematics, how to write well and be good citizens and so forth. This is the aspect to effectiveness that most schools understand because of the visibility of testing instruments in this process. But evidence of success (outcomes) must also be found in relation *all* school purposes, including those not typically tested. If schools expect students to be critical thinkers, to be capable of using wide range of communication skills, to be good at studying and taking notes, to be competent inquirers and cooperative individuals who can work effectively in groups, then the teacher's skills should be evaluated against these wider features as well. When was the last time a judgment of, say, an elementary school teacher's "effectiveness" at least partially accounted for whether her students loved to learn or loved to read recreationally? So, the idea is that effectiveness

must be accountable to a comprehensive construction of cognitive and non-cognitive effects or outcomes.

The second dimension to determining teacher effectiveness has to do with process concerns; that is, whether the decisions made by the teacher are professionally defensible, demonstrate receptivity to the nature of the learner and the values of society, and in the end, produce a learning experience that is attuned to the moving purposes of the school. If measured outcomes on a fairly limited range of tests (say, reading and mathematics achievement) are the only factors that calculate in the teacher effectiveness equation, then anyone who teaches to the test is all going to be viewed as a great teacher. We expect good teachers to not only positively affect key outcomes, but to do it in a way that speaks to a vital and dynamic learning experience. Thus, research claims that assert that 'effective' teachers are those who "use more low level questions and fewer high level ones," (presumably because such behaviors help to lift test scores), go face to face with a clearly articulated processual theory of teaching that would discount such assertions as unprofessional and largely out of alignment with what it means to be a responsive educator. All children, irrespective of evidence-based influences, should have teachers who demonstrate responsiveness to the nature of the learner, to the values of society and to some critical transmission of subject matter.

Conclusions

Best practices (and its evidence-based variants) have an easy and appealing logic to them. To argue against them puts the critic, at least at first glance, in the uncomfortable spot of favoring teaching that is not based on evidence. But one has to problematize the idea of evidence and the idea of what it means to be a good teacher to appreciate the point made against evidence-based instruction. If one admits that the construction of evidence in the evidence-based design is reducible to narrowly defined measures that yield conclusions against the aggregate, one could begin to pull at the thread that holds it all together. Teachers, especially those working with primary school age children, make instructional and pedagogical judgments about individuals, not groups (and certainly not statistically average groups), thereby putting into question whether anything like a scientifically-certified best method or what works strategy even exists in the educational situation. When a principal or a teacher might reasonable opt for an evidence-based practice or program (understanding that it might work best for most, not all), such an action admits the point, which is that individual students still need to be monitored and checked and the teacher needs to be prepared to make emergent adjustments in the education of the children. But when a principal or a teacher opts for evidence-based practice or programs truly believing that it is a best method or a

strategy certified by the government as effective, we could begin going down a dangerous road of stifling the discretionary judgment of the teacher, as she is reminded that the best method is at hand, and that no adjustment is necessary or desirable – that, in fact, fidelity to the best methods in the most desirable action. This could very well leave us with educators who are practicing methods and using programs that are fundamentally out of alignment with what teachers might otherwise view as best for the education of their charges. For this reason, schools across the country would be wise to be wary of anyone who peddles the idea of best practices.

References

Brusling, C (2005) *Evidence-based practice in teaching and teacher education.* Paper presented at the conference Professional Development of Teachers in a Lifelong Perspective, Copenhagen. Retrieved on January 2, 2008 from http://www.samford.edu/ctls/evidencebasedpracticeinteachingcb.pdf

Eisner, E (1998) *The educational imagination* (New York: Macmillan)

Dewey, J (1904). The relation of theory and practice to education In Society for the Study of Education 3rd yearbook, Part 1, *The relation of theory and practice to the education of teachers.* Bloomington, IL: Public School Publishing

Medley, D (1979) The effectiveness of teachers. In Peterson, P.L and Walberg, H.L. (eds.) *Research on teaching: Concept, findings and implications.* Berkeley, CA: McCutchen

Public Law 107-110, the *No Child Left Behind Act of 2001*, from http://www.ed.gov/policy (retrieved 12/18/07)

Rosenshine, B. V (1978) Content, Time and Direct Instruction In Peterson, P.L and Walberg, H.L. (eds.) *Research on teaching: Concept, findings and implications.* Berkeley, CA: McCutchen

Rules for a new kind of class: Interview with Howard Gardner (2007, September)IBWorldMagazine http://www.ibo.org/ibworld/sept07/rules_for_a_new_class.cfm (Retrieved 12/27/07)

Schwab, J.J (1983) *The practical 4: Something for Curriculum Professors to Do* Curriculum Inquiry 13(Fall):239-256

Slavin, R.E (2003, February) *A Reader's Guide to Scientifically-based Research* Educational Leadership 60(5): 12-16

Spillane, J.P and Miele, D.B (2007) *Evidence in Practice: The Framing of the Terrain* In Ross, P.A (ed). Evidence and Decision Making: the 106th yearbook of the National Society for the Study of Education (Malden, MA: Blackwell Publishing)

Tonelli, MR (2001). The limits of evidence-based medicine. *Respiratory Care* **46** (12): 1435–1440.