- White Paper -

Translating scientifically based evidence into student learning

Research analysis process and criteria

Instructional Focus and Causality

What local capacity-building strategy does your organization support to assist educators and schools at the local level in making decisions that support standards-based reform?

Read Diane Massell's *State strategies for building local capacity: Addressing the needs of standards-based reform* (CPRE Policy Briefs, No. RB-25) for an interesting 1998 description of needs—still as prevalent today as ever. Download (148 K) at: http://www.cpre.org/Publications/rb25.pdf

Regardless of your role, employing research that matters can help you to fulfill that role while supporting quality teaching and student learning.

Though many factors influence educational success, questions relating directly to instruction for student learning are the most important—and most difficult—to answer. Many organizations dodge the unavoidable certainty that what teachers and parents do with children, and how they do it, is the key determinant in student learning. They seek easier answers, often because they have never taught, and often because they are unable to make the difficult decisions that are required when attempting to make sense of findings that relate instructional interventions with progress in understanding by other human beings—our students. Classroom teachers and parents, and organizations that work to develop the real programs and products that constitute their principle resources for working with students on a daily basis, are not served well by such evasive efforts. Education programs and products must, when employed with students, result in desired learning objectives. To design and develop these programs and products so that their use, in the manner they prescribe, most closely results in that desired learning is, or *should be*, the goal of the resource provider. This requires prior work and honorable intent. It requires that we unabashedly pose questions that specifically address how certain instructional approaches and teaching strategies impact student learning. These are *causal* questions (e.g., does teacher use of strategy A cause potential student learning result B?). It requires that we consider a variety of information types—literature and experience-based viewpoint, studies employing correlational and pre- and post-designs, qualitative and observational findings, and empirical scientific evidence. All are useful, and the key is not *whether* one or another is used, but that reviews and discussions that guide further work recognize *which is which*, and do not attempt to obfuscate findings or viewpoints, or pass off one as another. The latter type of finding—scientific evidence—necessitates that studies on which we base our designs have employed a scientific approach. As well, it requires that program and product developers *judge* whether studies used, or conducted from the ground up, have indeed used a scientific approach. This means that criteria must be explicitly employed—not to rule out studies or sets of findings that are not strictly "scientific," but to rule out misinterpreting certain findings as evidence, or falsely claiming them as such.

Due to the scarcity of research evidence that directly targets teaching and learning, and the *translation* of findings into classroom practice, establishing and employing criteria means that developers and providers must first understand these criteria. They must also understand the reasons and limitations of the criteria, because it is a surety that decisions that are made must differ in some regard with each analysis and/or review conducted. The intent of this paper is to assist developers and providers with understanding these criteria, so that they may be employed more appropriately in each given circumstance, and so that claims that are made might more accurately reflect findings and the manner in which they have contributed or informed analyses and reviews upon which teachers and parents base their own instructional decisions.

As an education research and development firm involved in *translational education research* (e.g., lab to classroom, or translation of research into practice), Designed Instruction supports the efforts of organizations that strive to promote improved teaching and learning as a result of dedicated planning and research.

In the absence of thorough collections of evidence related to instruction, this paper describes certain processes we have initiated as part of an effort to bring clarity to illustrative practices, strategies and approaches utilized in the development and implementation of research-based education programs and instructional materials.

Analysis Process and Imperfection

Each analysis should involve an initial literature review of the instructional approach, its common use, issues, and associated variables. The information obtained should be synthesized, and from this synthesis a framework should be developed that allows analysts to isolate and focus on the most important aspect(s) of an instructional strategy or intervention. Note that the focus chosen depends on purpose—most typically in translational research, it involves determination of teaching strategies proven to be effective when used in certain situations and in a certain manner. Following determination of focus, analysts will need to conduct a detailed search for studies that provide evidence of effect on student learning, positive or negative. Based on certain criteria, developed from our own research and National Research Council and the U.S. Department of Education positions, we typically cull the studies from an initial count of 200-400 to around 10-15% that meet our criteria. The percent of studies used in the final analysis depends on the number of studies located that meet appropriate criteria, and the depth and breadth of the analysis desired. It is important to note at the outset that, regardless of the time or effort spent in locating quality studies, every useful and qualified study *will not* be located. Lack of inclusion of a particular study is not always a reflection on the quality or usefulness of the study. Though we believe it is important for a review or analysis of significance to reflect a certain number of studies that meet scientific (or sometimes quasi-scientific) criteria, we also believe in the intrinsic

The *No Child Left Behind Act of 2001* reauthorized the Elementary and Secondary Education Act and called for the use of "scientifically based research" as the foundation for education programs and student instruction. In February 2002, the Assistant Secretary for Elementary and Secondary Education hosted a seminar where education and science experts, including representatives from the National Research Council, discussed the meaning of scientifically based research across disciplines.

value of many literature-based, experience-based, and viewpoint-oriented works. These works simply should be recognized for what they are, used accordingly, and *not* included or referred to as a part of the scientific research evidence base. Where used in review, specifically note such use, and consider the use to shed relevant understanding on other more empirical findings. It is acceptable, respectful of the knowledge of so many highly experienced educators, and required in order to align findings with standards in order to blend the two into useful designs. Anything else would constitute a refusal to ponder and attempt to contextualize findings so that they can make sense, and prove useful, to teachers and parents.

Finally, it is also important to note that even of those studies considered to be “scientifically based,” very few will meet clear-cut gold-level standards for scientifically based research. We do not dismiss those studies that meet second-level, or quasi-scientific, standards in certain criteria, nor do we believe it is appropriate to use unrelated methodological reasons to omit findings that are honest and useful for informing or supplementing the evidence from another study or the combined resulting effect. We believe this would undermine the scientific veracity of the information obtained.

Scientific Research Criteria

In analyses, findings that are referred to as scientific evidence should be considered to be those that result from studies that exhibit scientific merit and relevance to the causal question being investigated. The following norms are ubiquitous in science—they apply to all sciences, including scientific education research. The norms are used in a practical, qualitative manner to measure the scientific validity of the studies that are located. All are interrelated, and dependencies (e.g., a norm being addressed in a study depending on whether or not another norm is addressed) changes from study to study, and within the context of each new analysis. Each study should be taken on a case-by-case basis, and determinations regarding inclusion should be based on a study’s usefulness both individually and as a whole in drawing a research-based set of findings in the context of the particular strategy and analysis being conducted. Selected studies and findings, either individually or as a contribution to the collective body, are:

> **Theory-based** - The research conducted should be concerned with addressing theories that can help to explain some aspect of teaching and learning, and at the outset the researcher should have an explicit or implicit theory or model of how a particular strategy or intervention is supposed to achieve its objectives. This improves the chance that the research is significant, relevant, and generates conjectures that can be more fully tested later. Both significance and relevance are sticky issues, and difficult to quantify or pin down in any sense. *Significance* means that the research directly informs the central problem or issue, builds on prior knowledge where possible, and represents a potential part of a larger research base that can inform instructional practice. *Relevance* refers to whether the study actually addresses what it was intended to address. For example, consider a study on the effect of computer 3D visualization on student ability to analyze data. Relevance is compromised if computers are used but without 3D visualization capability or software.

Controlled - Controls limit the degree to which outlying variables influence the results of a research study. There are many factors that influence learning. Controlling variables helps to ensure that the learning results achieved through use of a certain strategy can actually be attributed to that strategy. Clearly, control of studies related to learning, as with any study involving humans, can be very difficult. In studies reviewed, the primary characteristic analysts should hope to find is the recognition and attempt to control *confounders*—pre-existing characteristics of study participants that can negatively or positively bias research findings. Consider for example a hypothetical study regarding the effectiveness of government-mandated supplemental tutoring services. We should be able to presume that certain studies will compare performance of students who received these services with students that did not. Confounders that must be controlled in these studies might include past performance differences between these two groups of students, achievement levels and/or possibly even income levels of students who received these services compared to those who did not, or variations in parental oversight (or lack thereof) among students receiving services. These are a few of many possibilities, all of which deserve attention if the study is to result in scientific evidence. Usually, studies will use one of two primary methods for instituting control of confounding variables: *randomization*, and *selective statistical control*.

Randomization refers to assigning students to control and experimental groups with no attention to external or internal factors. If there are enough students involved across a diverse backdrop (and this is of paramount importance, as well as being subject to different interpretations of how much is “enough”), then by chance alone a study will involve groups that are comparable in every way except for the strategy being studied. This allows the effect of the instructional strategy to be uniquely isolated, and makes observed and measured differences in learning traceable to that strategy.

Randomization is the gold-level standard, because studies utilizing randomization require far fewer subjective interpretations when used to make causal inferences. However, practical limitations interfere. Consider the cost, time, and personnel required to conduct a large-scale random study. Consider the focused nature of research studies that inform specific instructional practices and approaches. Many times the two do not meet, so that an informed and useful review must often rely on studies that employ various lesser, though not insignificant or invalid, modes of control.

When a study does not, for a variety of possible reasons, utilize randomization, analysts should search for an explicit attempt to institute a modicum of *selective statistical control*. In our earlier example of supplemental tutoring services, we might find that a researcher separated student comparison groups by family demographics. Again, this is only one possibility. Almost invariably in the studies we encounter there is some uncertainty as to whether or not an important confounder was ignored. As a practical matter, and in light of the quality of other aspects of some of these studies, decisions must be made as to whether or not to include the study findings as a part of the analysis. In these cases, the most telling characteristic of whether or not a study is included is often the presence or lack of *critical evaluation* by the investigator. Cues are available, such as descriptions of disconfirming evidence and/or alternative explanations (e.g., the teachers in one group were simply more dedicated than those in another), how closely knit the project goals or ingredients were to the questions being investigated, how appropriate the measured variables in the study are for answering the question at hand, how generalizable the results of the study are to a certain population of students, how solid the control over time is (e.g., whether or not all students in the control stay in the control throughout the investigation), and so forth. Clearly, certain important characteristics are often juxtaposed. Analysts should seek investigations that are long-term enough to provide longitudinal data, but the longer a study persists the more chance there is that controls become biased. In summary, determinations are made, and our point is not to disclaim any investigator or the results of any investigation.

Measurable - Whether or not a study's results are considered measurable has very little to do with other studies that are located and included in an analysis. This norm relates more specifically to whether or not the researcher identified characteristics or indicators of learning that could be measured empirically, and whether or not he/she employed methods that permitted or facilitated a direct empirical investigation and measurement. In other words, was it measurable, and if so, were the methods used appropriate?

By *empirical* we mean the study should address a question that can be answered or at least informed through investigation or observation. It is important to note that some questions that cannot be measured empirically are essential, so that the acceptance of the research on this basis is not always an indication of the value of the research, but rather its "fit" with the overall context of an analysis.

Though methodology is a key feature of science, there are many legitimate methods that can be used, none of which individually determine the scientific veracity of the investigation. More important is the use of multiple methods that strengthen the certainty of the conclusions that are drawn. Some methods are, however, better than others for particular purposes. In each study, analysts should attempt to qualitatively determine how well particular methods used addressed the question at hand. Certain methods, while not as appropriate when considered in isolation, were very useful in allowing researchers to view questions from a variety of angles, search for disconfirming evidence, alternative explanations, and so forth, which relate to other norms.

Generalizable - As valid scientific inquiry rests in large part on how individual findings and results can be replicated in different times, places, and contexts, analysts must ask in a broad sense how the inferences drawn from each study—reliant as most are on limited numbers of participants, observations, and duration—generalize to a broader population or setting. As a whole, an attempt should be made to determine how often certain findings of a study are replicated by different studies and in different contexts. In all research, and perhaps in education more than any other, an intervention or strategy that works in one setting (classroom) may not work at all in another (e.g., a classroom just down the hall, or even in the same classroom with a different group of students). So, the conditions of the investigation are important. Studies in extremes (e.g., how well an approach works for highly awarded teacher) did not generally leave us with higher assurance that results would be replicated elsewhere.

Coherent - Though there is no single correct way to reason scientifically, analysts should search in each study for the NRC example of what John Dewey called a "warrant"—a scientific justification—for inferences and conclusions. Whether employed in a qualitative or quantitative sense, scientific reasoning is characterized by clear statement of assumptions, how and why evidence is considered relevant, and how much uncertainty exists in findings. This is a part of our culture of scientific inquiry, and it underscores once again the importance of the search on the part of the researcher for disconfirming evidence, the articulation of that evidence when found to exist, and how alternative explanations for what was observed were treated. Often this is assumed to have taken place when we find work that has, for example, been published in peer reviewed journals, presented at conferences, and the like. However, this does not necessarily determine the validity of a study. Therefore, though it is one possible indication that a study has been put to the test, it does not constitute a reason to summarily dismiss or include certain findings.

Designed Instruction, LLC is an education research and development firm dedicated to the improvement of teaching and student learning.

This paper is provided as a service to educators and organizations that are likewise dedicated to improving teaching and student learning through the use and implementation of quality research.

All research continually evolves, including the processes and criteria employed, as well as analyses and briefs based on the research. We always welcome, and regularly solicit, feedback and suggestions. As new information becomes available, its integration into our updated procedures and resulting documents is carefully considered.

Specific questions or suggestions? E-mail information@designedinstruction.com, or call toll free 1-888-276-2299

Designed Instruction
…providing research-based learning solutions

509 E. Park Street   www.designedinstruction.com   Sundown, Texas 79372