

Demonstrating “What Works” in Teacher Preparation Using “Gold Standard” Research Designs in Education

By Chris Sorensen, Sharon Smaldino and David Walker

“To control for environmental factors, all children born during this 18 month time frame would be assigned randomly to family units not a part of the original random sample.”

In recent years, new standards for education research have been defined. In the landmark legislation we all know as “No Child Left Behind,” there were more than 130 references to “evidence-based” or “research-based” education. In national publications and presentations, there is growing reference to a need for education to look at the Medical Model of research. In the document “Bringing Evidence Driven Progress to Education: A Recommended Strategy for the U.S. Department of Education” released in November 2002, an opening line reads ... “Our extraordinary inability to raise educational achievement stands in stark contrast to our remarkable progress in improving human health over the same time period” and the weakness of educational research is pointed to as the culprit. What is meant by “evidence-based” or “research-based” evidence is defined clearly in the Elementary and Secondary Education Act (ESEA):

- It employs systematic, empirical methods that draw on observation or experiment.
- It involves rigorous data analyses that are adequate to test stated hypotheses and justify conclusions drawn.
- It 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.
- It is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs or activities are assigned to different conditions 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.
- 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.
- Has been accepted by a peer reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective and scientific review. (It should be noted that a practitioner journal or education magazine is NOT the same as a peer-reviewed academic journal.)

Title II guidelines for Teacher Quality State Grants released in December 2002 include several interesting statements related to research in education:

- “Any agenda of research on teaching that ignores student achievement must seem a pointless intellectual exercise — educational researchers appear non-responsive to key constituencies and badly out of touch with the public mood.”

- “Findings from carefully controlled experimental evaluations must trump dogma. Expert judgments should be built on objective data that can be inspected by a broad audience rather than wishful thinking.”

These recent policies have implications for how we in education determine the effectiveness of our teacher preparation programs. We must make sure that we have “scientific evidence” to demonstrate that how we prepare teachers makes a difference in P-12 student achievement.

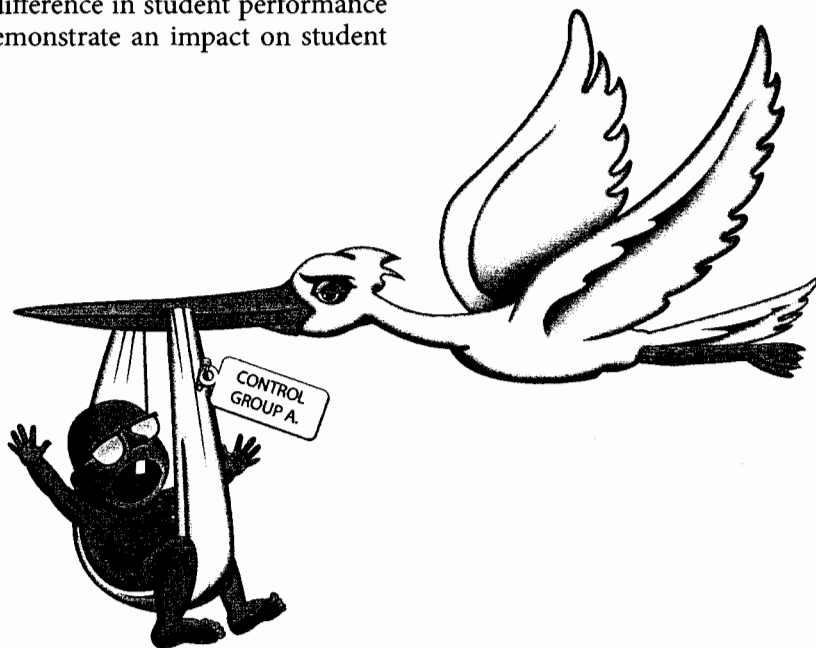
At a recent meeting in Washington D.C., we learned more about the “gold standard” for education research. In order to investigate causal relationships, we need to focus our research efforts on experimental designs. Random selection and random assignment are the keys to quality research.

At our university, we have a federal grant project targeted to improving teacher quality and teacher preparation programs. One of the questions the federal funders would like us to answer is whether the teacher preparation program in our university truly makes a difference in the quality of teachers prepared and, in the end, whether quality teaching makes a difference in student performance and achievement. The goal, of course, is to demonstrate an impact on student achievement.

As we pondered how to answer the question “Does this teacher preparation program lead to enhanced K-12 student achievement?” and in light of our new realization of the importance of a true experimental design to answer questions about “what works,” we began to design “the perfect study” that would truly give us an answer. For this perfect study, we knew we had to use both random selection and random assignment and that we had to have some controls in order to account for other variables that might influence student achievement. We came up with the following research design that we believe will allow us to determine whether our teacher preparation program impacts student achievement. We believe this design could be replicated in other teacher preparation programs in order to provide generalizable answers about “what works” in teacher education.

In looking at student achievement, the literature suggests that both genetics and environment can play a role in student performance. These variables can be controlled through a random assignment approach. In order to accomplish this, we believe we must take a longitudinal approach to the research. Literature suggests that a longitudinal approach is preferable in medical research, and thus we believe it should also be the preferred approach in education research. We have bounded our study to look at student performance in grades K-5. We recommend comparing four different teacher preparation programs, each with unique approaches to teacher education.

First, we would recommend that we randomly assign mating couples selected from a random sample of 5,000 heterogeneous individuals (half male and half female) representing various ethnic, socioeconomic and age groups and various academic backgrounds. This sample would be a national sample. The only criterion for inclusion in the population from which the sample is selected is that the individuals must be fertile and be able to provide medical documentation of that fact.



“During the course of the study, the teacher preparation programs at the participating institutions may not change their curriculum and must maintain the same teaching and clinical faculty.”

Once the 5,000 individuals have been selected, pairs of males and females will be randomly drawn and assigned until all members of the sample have been matched. Our hope would be that at least 2,000 children will be born within 18 months of the random assignment. We want to ensure that the children included in the study itself do not vary too much in age as rates of development might impact study findings. The use of randomly assigned mating couples should control for differences based on genetics.

To control for environmental factors, all children born during this 18 month time frame would be assigned randomly to family units not a part of the original random sample. To accomplish this, at least 2,000 family units would be selected randomly from across the nation, being sure to represent different family configurations (e.g., single parent, only child, nuclear family, extended family, same-sex couples, etc.), and also represent various ethnic, socioeconomic, regional and academic backgrounds. This random selection and assignment feature would control for environmental influences in the early development of the student sample. Once the children reach age five, 1,000 of them would be selected randomly to be assigned to the study schools.

For the purposes of this study, four K-5 schools will be selected randomly at the beginning of the study. One school from each continental time zone (i.e., Eastern, Central, Mountain and Pacific) will be selected. Because the literature seems to indicate that achievement problems are particularly noteworthy in high-poverty, urban districts, we propose that all four schools meet these criteria. Poverty and urban categorizations will be based on census data and definitions and school size for elementary buildings will be defined as having at least 200 students per building. While generalizability to other school settings may be limited, we believe controlling for size and type of school allows us to focus best on determining the impact of the teacher preparation program, the independent variable in which we are interested.

At the beginning of this study, four different teacher preparation programs will be selected to participate. Each program will be housed in a public university with enrollments of not less than 10,000 or more than 30,000, will represent a unique approach to teacher preparation and will be nationally accredited. One program each will be selected from each of the continental time zones. The graduates of the program must be able to be certified in the state in which the study school from that time zone is located. During the course of the study, the teacher preparation programs at the participating institutions may not change their curriculum and must maintain the same teaching and clinical faculty. Each teacher preparation program will be described in detail in the study report, including curriculum sequence, clinical preparation, student selection criteria, student rejection rates, and so on. Each school selected for the study may only hire teachers (and administrators) in their own time zone who graduated from the identified program (i.e., teachers graduating earlier would not be eligible). The school would have five years to relocate or dismiss existing teachers. The school administrators (principal and assistant principal if one exists) would also have to be graduates of the study university.

The variable of most interest, of course, is the learning of the students, as measured by standardized achievement tests aligned with learning standards. Students will be tested annually beginning in Kindergarten. Of most interest for our study will be student performance at the 6th grade level. Focusing on this longitudinal measure will help determine the longer term impact of the preparation programs. Students would remain in the study school for grades K-5 with annual assessments and determination of “value-added.” The assessment at 6th grade would determine which students were better prepared to succeed and to meet standards after leaving the study school.

In order to account for other factors (besides the teacher education program) that might impact student achievement, a number of factors in the schools would

be controlled carefully. First, teacher and administrator assignments would need to remain constant throughout the study. Second, all schools would use the same curriculum. Third, extra-curricular activities would be consistent across schools. Schools would need to have the same before and after school activities, the same student clubs and so on. Fourth, school resources would be equalized to ensure that no single school had access to resources not available to the other schools. Resources would include funding, technology, access to teacher professional development and the like. Finally, and perhaps most importantly, children would be housed in Living Centers where the environment would be tightly regulated to ensure that external, “home” factors were not influencing the study. Students would not be allowed to move or attend another school until 6th grade.

By implementing this “gold standard” study, we may be able to ascertain “scientifically” what really works in teacher preparation. With most other factors controlled, the differences in student achievement could truly be connected causally to the preparation of the teacher. This would certainly provide evidence which we could use to mandate practices in teacher education nationally. Of course, the practicality and cost of implementing this study, as well as ethical issues related particularly to efforts at controlling influences of other variables, must be examined to determine the viability of this approach to educational research.

Mark Twain once remarked “Always do right. This will gratify some people, and astonish the rest.” We are striving for this in educational research. Yet, there is one last issue that was forgotten. Our “perfect” study will have to be replicated. Ahh, the charge of *l'étalon d'or*!

Christine Sorensen is dean of the College of Education at Northern Illinois University. She teaches in the areas of research and evaluation and higher education. Her research has focused on the integration of technology in education and on the use of distance education. Prior to coming to NIU, she was an evaluation specialist at the Research Institute for Studies in Education at Iowa State University. Dr. Sorensen is the principal investigator for Project REAL, a federally funded Teacher Quality Enhancement project and is a co-author of Introduction to Education Research — 7th edition (Ary, Jacobs, Razeviah, & Sorensen) to be published this summer (2005).

Sharon Smaldino holds the position of Morgridge Endowed Professor of Teacher Education at Northern Illinois University. Dr. Smaldino has published extensively in the area of technology and education and is the author of several key books in the field. She is currently the president of AECT. At NIU, Dr. Smaldino directs school-university partnership activities for the College of Education and also directs Project REAL, a multi-million dollar Teacher Quality Enhancement project. Prior to coming to NIU in 2002, Dr. Smaldino was a professor of educational technology at the University of Northern Iowa.

David Walker is an assistant professor in the Department of Educational Technology, Research, and Assessment (ETRA) at Northern Illinois University. He teaches courses in statistics, research design and measurement. Some of his research interests include bootstrapping, effect sizes, predictive discriminant analysis, multiple linear regression, simulations and the creation of statistical computer programs. He presents research papers at conferences such as the American Educational Research Association (AERA) and the Mid-Western Educational Research Association (MWERA). He also consults as an evaluator on grants that look at P-12 issues.

“Of course, the practicality and cost of implementing this study, as well as ethical issues ... must be examined to determine the viability of this approach to educational research.”