Education Policy Analysis Archives

Volume 8 Number 53 **December 8, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Use of Logic in Educational Research and Policy Making

Rick Garlikov
Birmingham, Alabama (U.S.A.)

Abstract
While educational research is an empirical enterprise, there is significant place in it for logical reasoning and anecdotal evidence. An analysis of the article by Scott C. Bauer, "Should Achievement Tests be Used to Judge School Quality?" (*Education Policy Analysis Archives, 8*(46). Available: http://epaa.asu.edu/epaa/v8n46.html) is used to illustrate this point.

I want to use the following to help demonstrate the importance of logic, philosophy (particularly conceptual analysis), and insights based on anecdotal evidence, for educational research and policy making.

In "Should Achievement Tests be Used to Judge School Quality?" (*EPAA*, Vol. 8, Number 46) Scott C. Bauer stated the following:

> At the 1998 Annual Meeting of the Mid-South Educational Research Association, W. James Popham raised the following question: Is it appropriate to use norm-referenced tests to evaluate instructional

> quality? Specifically, he challenged participants to consider whether norm-referenced tests measure knowledge that is taught and learned in schools. Popham then invited researchers to participate with him in a study to answer the question: Should student scores on standardized achievement tests be used to evaluate instructional quality in local schools?
>
> In a subsequent paper, Popham (1999) laid out the basic argument that frames this study. While standardized achievement tests are useful tools to provide evidence about a specific students' mastery of knowledge and skills in certain content domains, "Employing standardized achievement tests to ascertain educational quality is like measuring temperature with a tablespoon" (p. 10). There are several difficulties with using aggregate measures from norm-referenced tests to judge the performance of a school. [Two of these are described, which I omit here.]
>
> [Third,] scores on standardized achievement tests may not be attributable to the instructional quality of a school. Student performance may be caused by any number of factors, including what's taught in schools, a student's native intelligence, and out-of-school learning opportunities that are heavily influenced by a students' home environment. Popham terms this last issue the problem of "confounded causality."
>
> Here we report the results of one of several local studies designed to provide empirical evidence to answer the question of whether student scores on standardized achievement tests represent reasonable measures of instructional quality.

This last sentence is only true if the term "reasonable" is understood to mean something like "credible to people who think about the issue in certain ways" or "credible to reasonable people who think about the issue in certain ways." It has to be understood in a way not dissimilar from the legal principle of considering "what a reasonable person would have believed or done in a similar situation" in order to assess the guilt or innocence of a defendant. This is because the study only actually surveys what people **believe** in regard to whether students who gave correct answers to individual standardized test questions were more likely to have been taught the information necessary to answer those test items in school or were more likely to have learned it elsewhere. The study did **not measure** whether students did learn the information in school or whether they learned it elsewhere, but whether teachers and parents **thought** students learned the information in school or learned it elsewhere.

Consider the following paragraph in Bauer's article:

> The notion that aggregate scores on standardized tests should serve as an indicator of school quality relies on an assumption of causality. The underlying logic is that the scores are predominantly caused by something the school does or has some control over. For this assumption to hold, at a minimum we must be willing to believe that student performance on standardized tests is related to school quality, that the tests measure the skills and abilities stressed in school programs, and that there are no antecedent factors that might otherwise explain aggregate student performance on the tests. If the

> data presented here are credible, the soundness of this assumption must be questioned. On average about half of the items on the rated test suffer from "confounded causality" on at least one of these criteria.

There is an ambiguity in the word "should", as he uses it, in the first sentence—the two meanings being (1) "should" in the political sense of whether policy ought to rely on standardized test scores to judge schools because people accept or believe that test items show direct causal correlations between the quality of school instruction and student test scores and thus, by extension, accept test scores of a measure of the efficacy of what is taught and learned in schools, (2) whether test items actually show direct causal correlations between school instruction and student test scores and thus serve as an actual measure of what is taught and learned in schools.

In the second sense it is not true that "For this assumption to hold [i.e., the assumption that scores are predominantly caused by something the school does or has some control over], at a minimum we must be willing to believe that student performance on standardized tests is related to school quality...." For the assumption to hold, what is necessary is that student performance on standardized test scores actually is related to school quality. Our beliefs about the accuracy of that statement have nothing to do with whether the assumption holds or not. We can believe it all we want, or disbelieve it all we want, and neither that belief nor that disbelief will make it true or false.

The proper conclusion is not that nearly half the items rated suffered from confounded causality, but that teachers and parents believed nearly half the items suffered from confounded causality.

The test for seeing **how much, if anything** of what is measured on standardized tests is actually taught in schools would require a very different kind of study—one which attempts either to find out precisely where students learned the information which they used to answer test items correctly, or at a minimum to find out whether students knew the information before it was taught in school or not, using some sort of pre-test/instruction/post-test differentiation methodology.

However, this latter would still only account for students learning the information prior to instruction. It would not account for students' learning the information during or after instruction, though **not because of** the instruction (alone). For example, it is a fairly common phenomenon for teachers to "teach" a principle that students do not understand, and that a parent or someone else then explains to the student in a way that the student comprehends it. Now it may be that the parent would not have done this without the teacher's introduction, but it is still then a joint teaching effort, not a result only of school instruction alone. And I suspect there is some evidence that in school districts where there is not such parent- or mentor-child interaction about school work, students do not learn it as well nor test as well. I also suspect that success on achievement tests, and academic or "grading" success in school in general comes in large part from parent or mentor interaction with school-initiated subject matter. The same argument could be given with regard to students' learning on their own—through reflection or additional study from other sources—material that was introduced in the classroom but that was not learned **in** the classroom nor **from** what the teacher (or textbook) said or did.

The point, however, is that **where and when** students have learned something **is** a **social science** kind of question, as is the question of where and when **what proportions** of students learn a particular item in school or elsewhere. And it is not dependent upon

where or when parents or teachers or anyone **thinks** students have learned something—unless the parent or teacher knows for sure. (The problem for the social scientists, however, in this latter case is ascertaining whether the parent does know for sure or not, because even if the parent is correct and does know, it is difficult for someone else to know the parent's claim is correct, particularly if the researcher or other third party was not present during the process.)

But now consider Popham's (or Bauer's, I can't tell which) claim: "Finally, scores on standardized achievement tests may not be attributable to the instructional quality of a school. Student performance may be caused by any number of factors, including what's taught in schools, a student's native intelligence, and out-of-school learning opportunities that are heavily influenced by a students' home environment."

If that is true, as it certainly seems to be since students do learn things, or figure out things, on their own or from others outside of school—things which sometimes are tested on standardized tests—**that** is alone sufficient to show that test scores **cannot** be reasonably attributable to instructional quality in schools alone. For if there are possible and reasonably likely other "confounding" or contributing causes of student success on standardized tests, then logic alone demands that test scores cannot **legitimately** be used to assess the quality of school instruction. Surveys about parent or teacher beliefs regarding this matter are unnecessary and logically irrelevant.

But that does not make this survey nor this paper unimportant. There are two things involved that are important. The first is that something may be politically popular even if it is not legitimate. So a survey of whether people think that standardized test scores reflect the quality of instruction in schools may be important to know for determining public policies (and news reporting policies) about using and/or reporting such assessments. If it turned out that the public did not have as much confidence in or concern about this form of assessment as legislators and newspapers seem to think they have, it might be politically feasible to get rid of these tests in a way that reasoning alone will not permit, because what is thought important to report in the news and what is thought necessary to legislate are often more dependent on what is believed to be desired by the public than on what reason might show is desirable or what evidence might show is false about public perceptions.

Second, this survey is interesting and useful as a teaching tool for the public, and in that regard is very important. For what Bauer has done is to show that people who look at individual test items are not confident about the significance of individual test item scores, and that **therefore they cannot be confident about the meaning or significance of aggregate scores**, and that, by extension, no one can be. It is one thing for someone to believe tests are significant without looking at and reflecting on the individual questions and the significance of each of them; it is quite another to believe that tests scores have significant meaning after examining the individual test questions and their likely significance. The survey was a way of getting people to do such an examination and to show them, and others, what happened when they did. For many people that is more convincing than logic alone, even if it should not logically be necessary.

I point out the above using the Bauer study because that study is not unique in educational research in regard to trying to demonstrate what is essentially a logical matter by use of empirical research. Further, it is not unique in educational research for researchers to draw logically unwarranted or unjustified conclusions from perfectly good data that they have collected. The point is that while logic and philosophy or conceptual analysis alone are often insufficient to provide knowledge about educational phenomena, they are both necessary in order to understand the significance of such data.

Moreover, they often show what data to seek. When Popham, or anyone, first realized that there logically could be confounded causality in regard to students' answering standardized test items correctly, that realization alone showed there was a problem that needed to be studied empirically in order to determine whether the logical possibility was the actual or likely or even systematic or overwhelming occurrence. But all too often in educational research and in educational policy-making, it is "empirical" research that is held to be all that is important, not logic nor anecdotal evidence nor insight based on anecdotal evidence. That seems to me to be a mistake because while logic and apparent single occurrences alone do not show what is happening systematically or statistically, they point out matters that either need to be studied empirically or they point to conceptual problems that may have to be addressed before empirical studies can be done. In some cases they also point out the actual futility of relying on a practice or policy that intuitively seems to be effective and that may even be traditional—such as determining the efficacy of schools by comparing (standardized) test scores. There are far more logical and conceptual matters involved in education and in educational research than is commonly believed or accepted. And I think it is a grave mistake to think that empirical studies alone are the proper or necessary way to do educational research and the only proper means to guide educational policy.

Reference

Bauer, S.C. (2000). Should Achievement Tests be Used to Judge School Quality? *Education Policy Analysis Archives, 8*(46). Available: http://epaa.asu.edu/epaa/v8n46.html

About the Author

Rick Garlikov
Rick Garlikov is a philosopher and photographer who resides in Birmingham, Alabama. He holds a graduate degree in Philosophy from the University of Michigan. He is the author of *The Meaning of Love*, of *Making the Most of Your University Courses: What to Expect Academically at College*, and of *Teaching About Thinking; Thinking About Teaching: Why Teaching "Facts" Is Not Enough*, an online book of essays about teaching for reasoning and understanding. Rick conducts introductory philosophy instruction via e-mail and offers a free "homework help" service that tries to help students (or parents) understand in greater depth the material with which they are having difficulty, so that they can then work through their assignments (or teach their children) on their own. These materials and services are available at http://www.garlikov.com.

EPAA Editorial Board

EPAA Spanish Language Editorial Board