The Obviousness of Social and Educational Research Results

N. L. Gage

Highly estimable writers have averred that well nigh all of the results of social and educational research are obvious, that is, could have been predicted without doing the research. To examine the justifiability of this allegation, one should examine its accord with actual research results. Thus, is it a "truism" that higher achievement comes about when students spend more time with the subject matter? That smaller groups are easier to control than larger groups? Do judges regard actual results as more obvious and statements of their opposites as nonobvious? Both the century-old research results of Joseph Mayer Rice and recent results throw light on these issues.

Educational Researcher, Vol. 20, No. 1, pp. 10–16

Is what we find out in social and educational research old hat, stale, platitudinous? Are the results of such research mere truisms that any intelligent person might know without going to the trouble of doing social or educational research?

The Importance of the Obviousness Question

The obviousness question has important ramifications. It can influence the motivation of anyone who is thinking about doing social or educational research. Why do research if you are not going to find anything new, anything not already known? Obviousness also relates to the justification of social science departments and schools of education in expecting or requiring their faculties and graduate students to do social and educational research. It also concerns government funding policies, such as those of the National Science Foundation and the National Institute of Mental Health that support social research, and those of the U.S. Department of Education, particularly the Office of Educational Research and Improvement, that support educational research. Foundations, school boards, state legislatures, and Congressional committees need to be convinced, before they put up the money, that social and educational research will produce something that any intelligent adult might not already know.

So, the issue of obviousness, apart from piquing our intellectual curiosity, has tremendous practical importance. Unless social and educational researchers face that issue, they may lack motivation to do research and lose societal support expressed in dollars.

The Charge of Obviousness

Does anyone really hold that social and educational research yields only the obvious? I begin with an old joke attributed to James T. Farrell, the novelist who became famous in the 1930s for Studs Lonigan. Farrell was quoted in those days as having defined a sociologist as someone who will spend $10,000 to discover the location of the nearest house of ill fame. He actually used a less polite term, and nowadays he would have said a quarter of a million dollars. I also remember a fellow graduate student who could always get a laugh by referring to the content of some of his textbooks as "unctuous elaborations of the obvious."

Schlesinger’s Critique

The first serious piece of writing that I know of that made the same charge appeared in 1949 in The Partisan Review. It was in a review by Arthur Schlesinger, Jr., of the two volumes of The American Soldier, which had just been published. The American Soldier was written by a group led by Samuel A. Stouffer, who later became a professor of sociology at Harvard. It reported on the work done by sociologists and other social scientists in surveying, with questionnaires and interviews, the attitudes of American soldiers during World War II. The first volume, subtitled "Adjustment During Army Life," dealt with soldiers’ attitudes during training, and the second, subtitled "Combat and Its Aftermath," dealt with soldiers’ attitudes while they were engaged with the enemy and risking their lives. As a young assistant professor, I found the two books impressive for their methodological thoroughness, sophisticated interpretation, and theoretical formulations of such concepts as "relative deprivation."

So I was taken aback after some months when I discovered a review of those two volumes by Arthur Schlesinger, Jr., the distinguished historian. Then a young professor at Harvard University, Schlesinger had just won a Pulitzer Prize for his Age of Jackson. Witty and vituperative, Schlesinger’s review also denounced what he considered the pretensions of social scientists. Schlesinger wrote:

"Does this kind of research yield anything new? ... [T]he answer... is easy. Most of the American Soldier is a ponderous demonstration in Newspeak of such facts as these: New recruits do not like noncoms; front-line troops resent rear-echelon troops; combat men manifest a high level of anxiety as compared to other soldiers; married privates are more likely than single privates to worry about their families back home. Indeed, one can find little in the 1,200 pages of text and the innumerable surveys which is not described more vividly and compactly and with far greater psychological insight, in a small book entitled Up Front by Bill Mauldin. What Mauldin may have missed will turn up in the pages of Ernie Pyle." (p. 854)

N. L. Gage is at the School of Education, Stanford University, Stanford, California 94305.
Lazarsfeld’s Examples

At about the same time as Schlesinger, Paul Lazarsfeld, a professor of sociology at Columbia University, also reviewed *The American Soldier*. Lazarsfeld (1949) was clearly aware of the same problem of obviousness. He wrote:

[I]t is hard to find a form of human behavior that has not already been observed somewhere. Consequently, if a study reports a prevailing regularity, many readers respond to it by thinking ‘of course, that is the way things are.’ Thus, from time to time, the argument is advanced that surveys only put into complicated form observations which are already obvious to everyone.

Understanding the origin of this point of view is of importance far beyond the limits of the present discussion. The reader may be helped in recognizing this attitude if he looks over a few statements which are typical of many survey findings and carefully observes his own reaction. A short list of these, with brief interpretive comments, will be given here in order to bring into sharper focus probable reactions of many readers.

1. Better educated men showed more psychoneurotic symptoms than those with less education. (The mental instability of the intellectual as compared to the more impressive psychology of the man-in-the-street has often been commented on.)

2. Men from rural backgrounds were usually in better spirits during their Army life than soldiers from city backgrounds. (After all, they are more accustomed to hardships.)

3. Southern soldiers were better able to stand the climate in the hot South Sea Islands than Northern soldiers. (Of course. Southerners are more accustomed to hot weather.)

4. White privates were more eager to become noncoms than Negroes. ([Because of their having been deprived of opportunity for so many years], the lack of ambition among Negroes was [quite understandable].)

5. Southern Negroes preferred Southern to Northern white officers [because Southerners were much more experienced in having interpersonal interactions with Negroes than Northern officers were].

6. As long as the fighting continued, men were more eager to be returned to the States than they were after the Germans surrendered [because during the fighting, soldiers were in danger of getting killed, but after the surrender there was no such danger].

(pp. 379–380)

Keppel’s Position

For a later sample of the worry about obviousness, we can turn to an essay by Frank Keppel, titled “The Education of Teachers,” which appeared in 1962 in a volume of talks on American education by American scholars that had been broadcast by radio to foreign audiences. Keppel had left the deanship of the Harvard Graduate School of Education to serve as U.S. Commissioner of Education under President Kennedy. As Commissioner he led the movement that resulted in the Elementary and Secondary Education Act of 1965, the first major effort in the U.S. to improve the education of children from low-income families. In his article, Keppel (1962) indicated that some people question the principles that have emerged from psychological studies of teaching and learning. Without committing himself as to whether he agreed, he summed up the critics’ arguments this way:

The efforts to use scientific methods to study human behavior seem to them [the critics] ridiculous if not impious. The result is a ponderous, pseudo-scientific language which takes ten pages to explain the obvious or to dilute the wisdom long ago learned in humanistic studies.... To build an art of teaching on the basis of the “behavioral sciences,” they suggest, is to build on sand.

(p. 91)

Conant’s Position

The very next year, obviousness was mentioned again, by another prestigious educator, namely, James Bryant Conant, who had been president of Harvard University for 20 years, and then the U.S. High Commissioner (and eventually the U.S. ambassador) in West Germany. During World War II, he had been a member of the highest scientific advisory committees, including the one that led to the production of the atom bomb. When he returned from Germany, he devoted himself almost exclusively to educational problems. In 1963, he published a book titled *The Education of American Teachers*, in which he reported on his studies of teacher education programs and schools—studies made through much interviewing, reading, and visiting. His book gained extremely wide and respectful attention. Yet, when I looked into it, as an educational psychologist, I couldn’t help being dismayed by Conant’s assertion that educational psychology largely gives us merely common-sense generalizations about human nature—generalizations that are “for the most part highly limited and unsystematized generalizations, which are the stock in trade of every day life for all sane people” (p. 133).

Phillips’s Critique

These references to obviousness take us only into the 1960s. Did the attacks disappear after that? Or are there more recent statements on the obviousness of educational and social research results? In 1985, a volume of papers appeared on the subject of instructional time, which had been central in a variety of formulations, such as John B. Carroll’s model of school learning, Benjamin Bloom’s mastery approach to teaching, and the concept of academic engaged time developed by Charles Fisher and David Berliner. All of these writers seemed to agree that the more time students spent in studying, practicing, and being engaged with the content or skills to be learned, the greater the related learning they achieved. The correlations between academic engaged time and achievement were not perfect, of course, because outside of the laboratory, correlations are never perfect, even in the natural sciences and certainly not in the social and behavioral sciences.

The subject of instructional time thus received a lot of attention in many articles and several books, including the edited volume, *Perspectives on Instructional Time*, to which the philosopher of the social sciences, Denis Phillips (1985), contributed a chapter entitled “The Uses and Abuses of Truisms.” Here Phillips first cited Hamlyn, also a philosopher, who had criticized the work of Piaget. Hamlyn had asked his readers to try to imagine a world in which Piaget’s main ideas were untrue:

a world where children mastered abstract and complex tasks before concrete and simple ones, for example. Such a world would differ crazily from our own, and one gets the sense that many of Piaget’s views are unsurprising and necessarily (if not trivially) true.

(p. 311)
Phillips then raised the same kind of question about the research on instructional time: “What sort of world would it be if children learned more the less time they spent on a subject? If achievement were not related to the time spent engaged on a topic?” (p. 311). So, just as with Piaget’s major findings, “one gets the sense that these findings [about instructional time] are almost necessarily (and perhaps even trivially) true” (p. 311). ‘Indeed, it suddenly seems strange to dress up these truisms as ‘findings’ ” (p. 312).

Phillips then went on to make a distinction between truisms and statements that are trivially true. “[T]he latter are, in effect, a subgroup of the former. A truism is a statement the truth of which is self-evident or obvious . . . whereas a trivially true statement is one that is true by virtue of the meaning of the terms involved (e.g., ‘All colored objects are colored,’ or ‘All bachelors are unmarried’)” (p. 312). He went on to say that “it is easier to keep a small group of children working on a task than it is a large group” is a truism, for it is obviously true, but it is not true by virtue of the meanings of the terms involved” (p. 312). Phillips also pointed out that:

truisms and statements that are trivially true are not thereby trivial. The terms truism and trivially true refer to the patentness of the truth of statements, whereas trivial refers to their degree of value or usefulness. The two do not automatically go together; many a statement the truth of which is far from obvious is of no practical use . . . and many truisms are vitally important and even theoretically significant (‘The sky is dark at night’ [this truism bears on the theory of the expanding universe]). (p. 313)

Furthermore,

truisms uncovered by researchers, then, are not necessarily trivial. But on the other hand truisms do not require research in order to be uncovered. Agencies would be wasting money if they awarded grants to researchers who wanted to determine if all bachelors in the United States were unmarried, or if the sky is dark at night, or if small groups are easier to control than large groups. (p. 313, emphasis added)

In Short

Let me summarize the argument so far. I have presented a series of opinions quite damaging to the notion that social and educational research yields results that would not already be known to any intelligent and thoughtful citizen. These opinions are hard to ignore. Extremely estimable people—Farrell, Schlesinger, Keppel, Conant, Lazarsfeld, and Phillips—all have made statements that might well give pause to any sensible person considering the pursuit of social and educational research or any organization being asked to part with money to support such research. I have presented these statements in chronological order extending from novelist James T. Farrell in the mid-1930s to philosopher Denis Phillips in the mid-1980s.

Empirical Examination of Obviousness

One noteworthy characteristic of all of these criticisms is that they were what might be called nonempirical or, at least, not systematically and formally empirical. Informal and personal, the appraisals were not made with any great specificity, detail, explicitness, or exactitude. Presumably, Schlesinger had not actually compared the statements of results reported in

The American Soldier with statements made by Bill Mauldin or Ernie Pyle. He did not perform a content analysis of the two kinds of reports about soldiers to show in any literal way that the sociologists’ statements of results had been anticipated by the insights of the cartoonist and the journalist. The same point can be made about what was said by Keppel and Conant: They did not go into any detail, or become at all specific, to support their allegations. However, the sociologist

I have presented a series of opinions quite damaging to the notion that social and educational research yields results that would not already be known to any intelligent and thoughtful citizen.

Lazarsfeld did go into detail and referred to specific results, namely, soldiers’ attitudes of various kinds. Phillips referred to specific findings about instructional time, or time on task, and also findings about size of group or class size.

Rice’s Studies

Now I should like to go back and look at some empirical efforts that seem to me to bear upon the whole issue of obviousness. I begin with what may be the first process-outcome study in the history of research on teaching. The results of this investigation were published by Joseph Mayer Rice (1897/1913) under the title “The Futility of the Spelling Grind.” Rice reported, after studying tests on 33,000 school children, that there was no correlation worth noticing between amount of time devoted to spelling homework and classwork and competence in spelling.

Rice’s evidence is still being cited in support of the argument that spelling competence results from “incidental” learning, rather than from any “systematic” teaching; that is, spelling is “caught” rather than “taught.” So far as instructional time or “academic engaged time” is concerned, the issue does not appear to be the open-and-shut case implied by Phillips (1985) when he asked, “What kind of world would it be if achievement were not related to the time spent engaged on a topic?” (p. 311). As Rice (1897/1913) put it, “concerning the amount of time devoted to spelling . . . an increase of time . . . is not rewarded by better results . . . . The results obtained by forty or fifty minutes’ daily instruction were not better than those obtained where not more than ten or fifteen minutes had been devoted to the subject” (pp. 86–87).

Apparently, showing a relationship between time on task and achievement was not as easy as falling off a log, as it should have been if the relationship between time-on-task and achievement were necessarily true, that is, a truism. At least in one subject matter, namely, spelling, the relationship between time-on-task and achievement was fragile, perhaps even nonexistent. So perhaps the relationship depended on
the subject matter. Perhaps other factors also made a difference. Things may be more complicated than we should expect if the relationship were a truism.

Similarly, if smaller groups were always easier to control, a relationship that Phillips assumed to be a truism, then they should show higher time-on-task and thus higher achievement. However, the trickiness of the relationship between class size and achievement is by now well established. Reducing class size from 40 to 20 does not improve achievement with any consistency at all. Glass (1987) reported that it required an “exhaustive and quantitative integration of the research” to refute well-nigh unanimous older assessments (e.g., Goodlad, 1960) that class size made no difference in achievement, student attention, and discipline. Even then Glass found that the relationship of class size to achievement appeared only probabilistically (in 111 of 160 instances, or 69%) when classes of approximately 18 and 28 pupils were compared. Moreover, the duration of the instruction made a big difference: the relationship was stronger in studies of pupils taught for more than 100 hours. In addition, the class size had to be reduced dramatically to make a major improvement: “Bringing about even a 10 percentile rank improvement in the average pupil’s achievement . . . may entail cutting class size (and, hence, increasing schooling costs) by a third to a half” (p. 544).

Alleging that a relationship (e.g., the size-of-group relationship to the ease of control) is a truism implies that it should always be found and that no exceptions should occur. Thus, all bachelors without exception are unmarried, all colored objects without exception are colored. By the same reasoning, if the group size-controllability relationship were a truism, all smaller groups should be easier to control than all larger groups. If the age-reasoning ability relationship were a truism, all older children should be capable of more abstract and valid reasoning than all younger children. But, of course, the last two examples are untrue. If a truism is “an undoubted or self-evident truth, especially one too obvious or unimportant for mention” (Webster’s New Collegiate Dictionary, 1979), then these relationships are not truisms because they are not always “undoubted” or “self-evident.”

Suppose we change the “truism” to a probabilistic statement (e.g., children tend to learn more, the more time they spend on a subject; time on task is positively but imperfectly correlated with achievement). Now the research aims to determine the strength of the tendency, or the magnitude of the positive correlation. Does the r equal .05, .25, .45, .65, or .85? It seems to be a truism that the size of the time on task versus achievement correlation depends on many factors: the reliability of the achievement measure, the variabilities of the two variables, perhaps the subject matter, and so on. Is the research to answer these important and specific practical questions still unnecessary?

Here may lie one key to the problem: To enhance the truism with the specifics that make it have value for theory and practice, the research does become necessary. Even if the broad generalization is a truism, the specifics of its actualization in human affairs—to determine the magnitude of the probability and the factors that affect that magnitude—require research. Even if “smaller groups tend to be more easily controlled” were a truism, we would ask, how much difference in group size is needed to produce a given difference in controllability? How do other factors—age and gender of group members, task difficulty, and the like—affect the difference in controllability resulting from changes in group size? Similar questions would apply to all the other seemingly truistic findings. Even if intelligent people could always (without any research) predict the direction (positive or negative) of a relationship between two variables, they could not predict its size and its contingencies without research-based knowledge.

Lazarsfeld’s Examples

Let us go back now to Lazarsfeld’s examples of obvious results from the World War II studies of The American Soldier. Recall his examples of the “obvious” conclusions from that study: better educated men showed more psychoneurotic symptoms; men from rural backgrounds were usually in better spirits than those from cities; Southern soldiers were better able than Northerners to stand the climate in the South Sea Islands; White privates were more eager to become noncoms than Black privates were; Southern Negroes preferred Southern to Northern White officers; and men were more eager to be returned to the States during the fighting than they were after the Germans surrendered.

Lazarsfeld (1949) asked, “Why, since they are so obvious, is so much money given to establish such findings?” However, he then revealed that

Everyone of these statements is the direct opposite of what was actually found. Poorly educated soldiers were more neurotic than those with high educations; Southerners showed no greater ability than Northerners to adjust to a tropical climate; Negroes were more eager for promotion than whites, and so on . . . If we had mentioned the actual results of the investigation first, the reader would have labelled these “obvious” also. Obviously something is wrong with the entire argument of obviousness. It should really be turned on its head. Since every kind of human reaction is conceivable, it is of great importance to know which reactions actually occur most frequently and under what conditions . . . (p. 380)

Lazarsfeld’s rhetorical ploy has always impressed me as fairly unsettling for those who make the allegations of obviousness, but its force depends on whether we are willing to grant him his assumption that we accepted the first version of the research results as valid, so that he could then state his with his second presentation, which gave the true findings: the results that were actually obtained. It might be argued that Lazarsfeld’s assumption was unwarranted and that most of us would not have believed that first set of statements that he later revealed were spurious.

The Mischels’ Study

So I took notice when I heard about investigations that made no assumptions of the kind that Lazarsfeld’s exercise required. The first of these (Mischel, 1981; Mischel & Mischel, 1979) consisted of giving fourth- and sixth-grade children (Ns = 38 and 49, respectively) items presenting psychological principles stated in both their actual form and the opposite of the actual forms. For example, the first item dealt with the finding by Solomon Asch that college students would respond contrarily to the evidence of their senses about which of three lines had the same length as a comparison line when the students first heard four other students (confederates of the investigator) misidentify the same-length line. The second item concerned Harry Nelson’s finding that the same water temperature feels cooler on a hot day than on a cool
day. In all, there were 17 such items, some of which were presented to only one of the two grade-level groups. The children circled the one of the two to four choices that they thought described what would happen in each situation.

Of the 29 opportunities for either the fourth graders or the sixth graders to select the actual research result to a statistically significant degree, the groups did so on 19, or 66%. One group or the other was wrong to a statistically significant degree on five opportunities, and there was no statistically significant correctness or incorrectness on 5 opportunities. Clearly, the children had substantial success, but far from the perfect record that would support the allegation of almost universal obviousness.

But these were only children. What about college students and adults? And what happens when the research results are presented as flat statements rather than as multiple-choice items requiring the selection of the actual result from two or more alternatives?

Baratz’s Study

Baratz (1983) selected 16 social research findings from various studies, and then did an experiment. She manipulated, for each of the findings, whether the statement concerning that finding was the true finding or the opposite of the true finding. She also presented each finding, either the true one or the opposite one, with or without an explanation of the finding. That second manipulation was intended to “explore the possibility that adding explanations to the findings may render the findings more obvious” (p. 20). Thus, each of her subjects—85 male and female undergraduates enrolled in introductory psychology at Stanford University—evaluated 16 findings: four statements with a true finding plus explanation, four statements with the opposite finding plus explanation, four statements with a true finding without explanation, and four statements of an opposite finding without an explanation. Each finding was presented in the same format: first, the question addressed by the study, such as “a study sought to determine whether people spend a larger proportion of their income during prosperous times or during a recession.” And for this study the reported finding was “In prosperous times people spend a larger proportion of their income than during a recession.” The statement of the opposite finding differed from that of the true finding only in the order of the critical terms, and half of the findings were followed at the time by a short explanation, which was presented as the “explanation given by our subject.”

Here are two sample pairs of the true and opposite findings used by Baratz in her experiment: “People who go to church infrequently” versus “People who go to church infrequently tend to have more children than people who go to church regularly” and “Single women express more distress over their unmarried status than single men do” versus “Single men express more distress over their unmarried status than single women do.”

For each of the 16 findings presented to each student, the students were asked how readily predictable or obvious the finding was and were instructed to choose one of the responses on the following four-point scale:

1. I am certain that I would have predicted the result obtained rather than the opposite result.
2. I think that I would have predicted the result obtained rather than the opposite result, but I am not certain.
3. I think that I would have predicted the opposite to the obtained result, but I am not certain.
4. I am certain that I would have predicted the opposite to the obtained result.

The subjects were asked to express their “initial impressions of the relevant findings, i.e., the kind of impression that you might form if you read a brief article about the research in your daily newspaper” (p. 25).

In a summary table, Baratz presented the mean percentage of subjects who marked either “I am certain that I would have predicted the reported outcome” or “I think I would have predicted the reported outcome” for pairs of opposite findings. When the reported outcome was “A,” 80% of her students claimed they would have predicted that outcome. When the reported outcome was “B,” 66% of her subjects claimed they would have predicted that outcome. Thus, as Baratz put it, “It is clear that findings that contradict each other were both retrospectively judged ‘obvious’ . . . These results show clearly that reading a result made that result appear obvious. No matter which result was presented, the majority of the subjects thought that they would have predicted it” (p. 26).

I considered Baratz’s experiment and her findings to be persuasive. They seemed to provide evidence against the argument that social research yields only obvious findings. Her results indicated that intelligent people, namely, Stanford undergraduates, tend to regard any result they read, whether it is the true one or the opposite of the true one, as obvious. This tendency to say results are obvious was, of course, only a tendency: not all of her subjects followed that tendency, but it was a majority tendency.

Wong’s Study

Baratz’s research on obviousness dealt with results from a fairly wide range of the social sciences, but I had been focusing on research on teaching and particularly on one area within that field: process-outcome research. That kind of research seeks relationships between classroom processes (what teachers and students do or what goes on in the classroom) and outcomes (what students acquire by way of knowledge, understanding, attitude, appreciation, skill, etc.). Would such research results elicit obvious reactions similar to those obtained by Baratz?

A few years ago, Lily Wong, a Stanford graduate student from Singapore, replicated and extended Baratz’s experiments, but with findings from process-outcome research on teaching. Wong chose her respondents from four different categories of persons who differed on the dimension of how
much they might be expected to know about classroom teaching. At the low end of that dimension were undergraduates in engineering; next, undergraduates majoring in psychology; next, teacher trainees; and at the high end, experienced teachers. Each of these four groups of respondents was sampled both from Singaporeans and from Americans residing either at Stanford University or in the neighboring area. In total, Wong used 862 Singaporeans and 353 Americans. For the research findings, she used 12 statements based on results of process-outcome research carried out in the elementary grades, results that had been cited in the third edition of the Handbook of Research on Teaching (Wittrock, 1986) and in textbooks of educational psychology. Her items came from the results of research by Anderson, Evertson, and Brophy; Brophy and Evertson; Good and Grouws; Soar and Soar; and Stallings and Kaskowitz. Here is the first of her 12 items: “When first-grade teachers work on reading with a small group of children, some attend closely to just the children in the small group, whereas others monitor children’s activities throughout the classroom. The class’s reading achievement is higher when teachers monitor the entire classroom” versus “...when teachers attend to just the children in the small group.” Here is the second item: “When first-grade teachers work on reading with a small group of children, some call on the children in a fixed order, whereas others call on children in a random order. Reading achievement is higher when children are called on in a fixed order” versus “...when children are called on in a random order.”

Wong had five forms of questionnaires: Form A, Forms B1 and B2, and Forms C1 and C2. Subjects completing Form A had to select in each item the true finding between two options—one stating an actual finding of research on teaching at the primary-grade level and the other stating the opposite of the actual finding. The subject then rated the chosen statement on a 4-point scale from 1, “extremely obvious” to 4, “extremely unobvious.”

Subjects completing Forms B1 or Form B2 were required to rate the obviousness of each of 12 single statements presented as actual research findings. In fact, 6 were true findings and 6 were the opposite of true findings. Each of the 24 statements from Form A thus appeared in either Form B1 or Form B2.

Form C subjects were given the same purported findings as Form B subjects, but in Form C, each statement was accompanied by a possible explanation. Subjects in Form C had to rate not only the obviousness of the findings but also the clarity of the explanations.

Wong’s results on Form A showed that her respondents chose both actual findings and opposite findings. On 4 of the 12 items, her subjects chose the actual finding more often (see p. 37), but on the other 8, they chose the false finding more often. The r between percentage choosing a finding and the mean obviousness rating of the finding was .66. The respondents to Forms B and C rated about half of the opposite findings as obvious. Wong concluded that

Judging by the smaller proportions of respondents choosing the actual findings as the real findings, and the mean rating of obviousness on the presented (both actual and opposite) finding statements, we can say reasonably that people can not distinguish true findings from their opposites. (p. 86)

The Singaporeans rated most of the items as more obvious than the American subjects did in all conditions. There were few gender differences in the average responses to the various forms. Teachers were no more accurate, on the average, than the other groups in the selection of true findings: “In the rating of obviousness of items, knowledge and experience [in teaching] were found to have some significant effect on several items. This does not mean that teachers and trainees rated true findings more obvious or opposite findings less obvious than the psychology undergraduates and the engineering undergraduates” (Wong, 1987, p. 87).

Wong concluded that her results “clearly confirmed the idea that knowledge of outcome increases the feeling of obviousness. Thus, when people claim to have known it all along when an event is reported to them, their claim is often not warranted” (p. 88).

Where the Issue Stands

From the work of Baratz and Wong we can conclude that the feeling that a research result is obvious is untrustworthy. People tend to regard as obvious almost any reasonable statement made about human behavior. A recent example comes from the Arizona Daily Star of March 8, 1988, in an article about the booklet entitled What Works, compiled by the U.S. Department of Education. The booklet contains brief discussions, with references to the research, of 41 research findings considered potentially helpful to schools and teachers. The headline read, “Restating the Obvious.”

My most recent example comes from the June 1990 issue of The Atlantic (Murphy, 1990): “A recent survey (by me) of recent social-science findings... turned up no ideas or conclusions that can’t be found in Bartlett’s or any other encyclopedia of quotations” (p. 22).

As suggested by an anonymous referee for this article, the results of Baratz and Wong are consistent with the conclusions of Nisbett and Wilson (1977): “[T]here may be little or no direct introspective access to higher order cognitive processes” (p. 231). Thus the cognitive processes that lead one to regard a research result as obvious are probably nonveridical unless, as Ericsson and Simon (1980) argued, the response is based on (a) short-term memory leading to verbalization of information that (b) would have been attended to even without the instructions given. It is questionable whether judging the obviousness of research results always meets these requirements.

The same reviewer also suggested that these results do not belie the fact that most adults’ generalizations about human interactions are at least functional. I agree; otherwise human society would be impossible.

Another issue arose in a conversation between Robert D. Hess and me. Upon being apprised of judges’ tendency to regard as obvious both actual research results and their opposites, Hess asked about the frequency with which the results had been confirmed through replications. His question calls for research in which the “obviousness” of research results frequently confirmed with high consistency would be compared with that of research relationships frequently studied with results of only low consistency. Examples of both high-consistency and low-consistency results can be found in the synthesis of results of research on teaching by Walberg (1986). His Table 7.2 (pp. 218–219) contains results whose “percentage positive” across replications ranges from very low (where 50% is completely inconsistent) to very high (where 0% and 100% are completely consistent).
An investigator could administer questionnaires similar to those of Baratz (1983) and Wong (1987), but using items representing both (a) frequently studied with highly consistent results and (b) frequently studied with highly inconsistent results. It would then be possible to determine the difference, if any, in the mean obviousness rating of these two types of research results. It may turn out that only items of Type b would be rated obvious in both their actual and opposite forms. A frequently replicated and highly consistent result—for example, the "result" that auto drivers in England stay to the left side of the road whereas auto drivers in the United States stay to the right side of the road—will almost certainly be rated highly obvious in its actual form and highly nonobvious in its opposite form. Here the requisite knowledge is widely possessed, and the "obvious" reaction will not occur. Much depends on the relationship between the content of the research result and the background knowledge of the judge of the result's obviousness. Both Baratz and Wong may have studied results whose relationship to their judges' background knowledge was tenuous. Even Wong's experienced teachers, who probably had never thought about or encountered the phenomena dealt with in the research results used by Wong, had too little background knowledge to be able to detect the nonobviousness of the opposite-to-actual results.

Thus, the obvious reaction may be hypothesized to occur only when the judge's background knowledge in relation to the judged research result is weak. If the hypothesis is borne out, the question might be raised. How does a representative sample of social and educational research results fare, as to their obviousness in actual and opposite forms, when presented to a representative sample of the persons who might be expected to encounter or be concerned with those results? That is, research on obviousness now needs to be aimed at maximal external validity, or the degree to which the obviousness research is relevant to real life.

The issue joined by Schlesinger when he attacked students of human affairs who use scientific methods has its roots in the old controversy that C.P. Snow (1964) examined later in The Two Cultures: And a Second Look. Snow was concerned with the mutual disregard and disrespect of natural scientists and scholars in the humanities. Snow regretted this condition, but it still exists. Schlesinger's denunciation of social research reflected what Karl Popper called the antinaturalist position: the position that the scientific method useful for studying the natural world is inappropriate for the study of human affairs. The response of Paul Lazarsfeld reflects the position, held by Karl Popper and many others, that scientific method is appropriate for the study of human affairs.

Scientific method need not be used, in my opinion, only for the construction of a social science—where such a science is defined as a network of laws that will hold over whole eras and in many different cultural contexts, just as the laws of mechanics hold in different historical periods and in contexts as different as planetary motion and the motion of a pendulum. Rather, scientific method can be used for what Popper called "piecemeal social engineering," a more modest enterprise aimed at improving human affairs by applying scientific methods to the development and evaluation of new "treatments"—in education, in social welfare projects, or in fighting against drugs.

I have speculated (Gage, 1989) that people gravitate toward one or the other of Snow's two cultures—toward science (natural or social) or toward humanistic insight and sensibility—because their upbringing and intellectual experience have inclined them toward one or the other. The wars between the several paradigms in social and educational research may result from temperamentally different (i.e., not entirely rational) intellectual predilections, often developed during the secondary school years. If so, improved education may someday produce scholars and educational researchers who experience no conflict between their scientific and humanistic orientations.

In any case, the allegation of obviousness may now be countered with the research result that people tend to regard even contradictory research results as obvious. Perhaps even that result will henceforth be regarded as obvious.

Note
This article is based in part on the Maycie K. Southall lecture at George Peabody College, Vanderbilt University, on February 27, 1990.

I am grateful to my daughter, Sarah Gage, for calling the Murphy (1990) article to my attention.

References