DOCUMENT RESUME

ED 227 735 HE 015 972

AUTHOR Wagner, Paul A.
TITLE Classes of Legitimate Evidence for Identifying Effective Teaching.
PUB DATE [83]
NOTE 15p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Codes of Ethics; *Faculty Evaluation; Higher Education; *Peer Evaluation; *Student Evaluation of Teacher Performance; *Teacher Effectiveness

ABSTRACT

A criterion for selecting sources of evidence to evaluate effective teaching is described. It is suggested that teaching effectiveness is not measured solely in terms of cognitive change in students but in the extent to which academics practice teaching in accordance with the moral dictates of the profession. In developing a teacher effectiveness evaluation criterion, it is important that judgments be as objective and fair as possible, even though the selection of attributes of teacher effectiveness necessary reflects the biases of the evaluator. It is argued that there is no rationale for soliciting student responses as a measure of effective teaching, since a person who knows nothing about a specific set of skills and information as well as the larger discipline is in no position to comment on the academic merit of students' acquired knowledge of such materials from a given teacher. On the other hand, by examining student examinations, discipline specialists are in a position to make an educated guess about the extent of student learning resulting from course participation. Other faculty members are also good sources of judgment about the sufficiency of topics covered in the course. It is unlikely that students can make informed judgments about whether an instructor has effectively taught in a morally responsible way. (SW)

Classes of Legitimate Evidence for Identifying Effective Teaching

by Paul A. Wagner, University of Houston at Clear Lake City

Empirical research into university teaching, like any other empirical research, is wholly dependent upon theory for its initiation, practice and successful completion. It is theory which governs which classes of phenomena are relevant to a particular study and it is theory which determines which observations are to count as evidence in support of a specific claim.

Researchers in the physical sciences and more recently the social sciences have long recognized the failure of operationalism to give us a theory--free basis for scientific study. No observation is directly accessible to our consideration without the benefit of theory. Rather, as Norwood Hanson and many others have shown, every observation, at every level of scientific study is ultimately theory--laden in some important sense. The attempt to guarantee scientific objectivity by the use of operational definitions has proved itself to be a dismal failure. Sophisticated and conscientious researchers in all areas of scientific practice today recognize the continued priority of theory--making over data collecting. Theory-making is logically prior to data--collecting because theoretical concerns dictate how the researcher sees the world as well as how such observations are to be organized in refuting or confirming a specific

hypothesis.[2] Good theory not only prescribes and limits speculations to be used as future research hypotheses, but theory articulates a criterion for making observations and for identifying some observations as relevant to the evidentiary claims of a certain hypothesis while denying to other observations any role at all in such determinations.[3] In studying effectiveness in university teaching due attention must be given to the theoretical determinants governing the nature of such research before a researcher is in a position to make any claims based on the results of such research. Specifically, a theory of effective teaching must be sufficiently developed in order that a taxonomy can be developed for distinguishing relevant from irrelevant observations. Too often in the past the important theoretical tasks in this area have been ignored. Instead, many researchers have employed the philosophy of operationalism as a license permitting them to associate personally favored sets of practices with attributes the researchers have proclaimed are characteristic of effective teaching. Since the evaluation of effective teaching is so important to students and recently to the careers of practicing teachers, we can no longer allow such research to proceed in such a glib and arbitrary manner. In what follows I will attempt to sketch a criterion for determining what classes of observations are relevant to making claims about effective teaching. In general, I will argue that there is a conceptual link existing between such things as effective teaching and peer evaluation, the moral

sensitivity of the instructor and the instructor's recognition of being a professional. On the other hand, there exists no rationale whatsoever for soliciting student responses as a measure of effective teaching.

In determining what counts as relevant evidence for determining teaching effectiveness some researchers have detailed many attributes which they speculate are relevant. Others, even such a meticulous thinker as Michael Scrivens, [4] have at times been almost glib on this matter. These latter researchers claim that when all is said and done the one observation that really matters pertains to the amount of material learned by the students. If one were to take this position seriously then when evaluating effective university teaching, one could ignore whether or not what the students learned was correct, important, useful and so on. The only important question is simply, "did they learn a lot of whatever they are studying?" Similarly, if one were concerned solely with the amount of material acquired by the student then one could also ignore any practice an instructor might employ when getting students to learn since we are only concerned with the amount of material acquired by the student and not with how it is acquired.

However, in the real world of university teaching, the fact of the matter is we are not concerned <u>solely</u> with the amount of material learned by each individual student. In the real world of university teaching we are concerned for

example, that no faculty member breaks the law when facilitating the student's learning. Similarly, we are, or ought to be, concerned with the moral consequences of different teaching techniques and the meritoriousness of the material the student is being taught. For example, in the case of the former we condemn, or ought to condemn, an instructor's attempt to encourage students to behave like Neitzchean supermen and commit crimes such as that of Loeb and Leopold, the two infamous student-murderers of the 1920's from the University of Chicago. Even if it could be shown that a particular student learned more as a direct result of commiting the crime than he or she might otherwise learn is not sufficient justification for allowing certain events to occur. Similarly, in the case of the latter, we condemn, or ought to condemn, any course in which the instructor sets out to teach inaccurate historical information so that the students come to accept the same political doctrines as the professor.

Now there are no doubt those among us who would argue that such concerns are extraneous to the evaluation of teaching effectiveness. Such people generally argue that teaching effectiveness applies only to the cognitive change that occurs in a student as a result of a curricular experience engineered by an instructor. However, it is simply not the case that the evaluation of university teaching effectiveness occurs in a vacuum and is concerned solely with cognitive changes in the student that result from the teacher's

efforts. For example, in a class that strikes a student as exceedingly tiresome, the ingenious student may conjure up countless ways of adapting to his boredom. Each way represents something that the student learned directly or indirectly as a result of his experience within that particular class. Nevertheless, no teaching effectiveness instrument with which I am familiar, makes any attempt to detect such changes in the student's conceptual apparatus. In fact, the only way that all such changes could be detected would be if something akin to engram formation truly takes place and if some neurophysiological investigation could detect a statistically significant number of such changes. I am not arguing that no evaluation of a change in the student's cognitive structure in response to teacher activity can be made, I am merely making the rather obvious point that whatever determinations we do make will depend on our own conceptions of what we deem is important in the academic environment and this is as inevitable as it is proper.

As noted above, we could not measure pure and simple "cognitive change" even if we wanted to. But even if we could, I do not think that would settle the problem of evaluating university teaching. When discussing the "real world" matter of evaluating university teaching we are not discussing a relatively sterile concept such as engram formation. As noted above, any evaluation of university teaching belies the investigator's interest in detecting those sorts of conceptual changes which he or she thinks should result

from a successful academic experience. And, as mentioned above, such normative concerns are not only inevitable but they are proper as well. Only the leaders within an academic discipline are in a position to know what changes in an individual's conceptual apparatus ought generally to result from a particular experience. And, only those who are members of a community are in a position to identify moral principles which limit and promote various practices within that same community. University teachers are professionals and as such they are members of a community. In the real world in which the evaluation of university teaching takes place, academics, as professionals, are, or ought to be, concerned with how colleagues attempt to bring about changes in the conceptual structure of students. On the one hand, we argue that instructors should be guaranteed considerable latitude in the pedagogical practices they employ when teaching. This guarantee is an essential aspect of the principle of academic freedom. On the other hand, academics recognize that there are limits even to academic freedom. Thus, faculty members would generally condemn a colleague who distrubuted heroin to his students so they could experience the high of an addict or required students to attend a sexual orgy so they could become more familiar with the "feel" of an orgy activity. In short, in the real world of the university, teaching effectiveness is not measured solely in terms of cognitive change in students but in the extent to which academics practice

their teaching craft in accord with the moral dictates of the profession.

Prior to constructing a criterion for establishing classes of evidence relevant to the assessment of teaching effectiveness, the following points ought to be noted:

1. Judgments regarding teaching effectiveness ought to be as objective as possible.

2. Judgments regarding teaching effectiveness ought to be done as fairly as possible.

3. Teaching effectiveness is largely (though not solely) a matter of getting students to acquire novel information and skills of an academically important kind.

4. Judgments about the attributes regarding teaching effectiveness are necessarily a product of the biases of individual researchers.

5. University teaching is a craft practiced by a group of professionals and is subject to the moral restrictions of that specialized community.

In lieu of the above restrictions limitations on the assessment of university teaching effectiveness seem evident. The first two points above dictate that any responses from members of the university community regarding the teaching effectiveness of an individual instructor be done by a

knowledgeable, professionally competent and fair--minded judge. Only a knowledgeable and professionally competent person is in a position to judge the academic merit of information and skills acquired by students as a result of a particular curricular experience. Obviously, a person who knows nothing about a specific set of skills and information is unable to judge if some agent X has successfully acquired such skills and information. Similarly, a person who knows nothing about a specific set of skills and information as well as the larger discipline of which such skills and information are said to be a part, is in no position to comment on the academic merit of a student's learning such materials. For example, students presumably take college courses because they are innocent of the material such courses are supposed to contain. At the end of the semester we have no reason to believe that they are any more knowledgeable of the intended material than was presented during the course of the semester. Consequently, students are in no position to comment on the amount of material learned in a particular course or its academic relevance. On the other hand there is reason to believe that other specialists in the discipline have sufficient perspective to pass judgment upon the merit and extensiveness of the material presented in a particular course. In addition, by examining student examinations, discipline--specific specialists are in a position to make an educated guess as to the extensiveness of the student learning which resulted from participation in the course. This is not to

say that discipline specific specialists can make fully informed and otherwise adequate assessments of the nature and extent to which specific material was learned in a course, it is only to note that there is absolutely no reason in theory to conclude that students are in such a position.

The principles of fairness and objectivity noted above similarly preclude any attempt to solicit from students any responses regarding the personality characteristics contributing to a teacher's style of instruction. College students like any other adult, have certain preferences and biases that make them more attracted as individuals to one person than to another. Now it is often the case that one person can find many attractive characteristics in another person and yet not learn much of anything from that person. In fact, it may well be a person's engaging personality and demeanor which mitigates against a student's learning from that person. In any case, the personality attributes a student may be inclined to rate high when judging a particular instructor may have nothing at all to do with that instructor's teaching effectiveness. In addition, keep in mind that instruments including reference to certain personality attributes were constructed by an evaluation specialist with his or her own set of biases. Some evaluation specialists are quick to appeal to the fact that the attributes included on their instrument are supported by other research studies. Unfortunately, there are two factors which make such claims of trivial importance at best. First, it is quite easy at

this point in time to find some study somewhere that will endorse the instrument maker's selection of a personality attribute for inclusion in the new instrument. Consequently, in choosing one set of personality attributes from among the many attributes that are supported by research findings in the literature is a direct consequence of previous research bias. There just seems to be no way around this charge. Second, research findings regarding personality attributes are supposedly validated in a technical sense if they produce results commensurate with the results of an earlier study. Such a procedure is subject to devastating criticism employing a reductio ad absurdum argument. If each study is technically validated because its results are commensurate with the results of an earlier study then each subsequent study is fettered by the biases of the earliest study cited.

Again, the point to be gleaned from the discussion immediately above is not that there can be no evaluation of personality attributes likely to contribute to effective teaching, but that the extent and intensiveness of biases affecting such evaluations are compounded and reach an intolerable level of capriciousness when student responses are solicited for the purpose of judging teaching effectiveness. At this point in time there is little to recommend any source of observation as evidence for determing if instructor Jones has an optimal array of personality characteristics for teaching students X, Y and Z some specific subject matter.

Finally, as noted early on in the discussion above, university teaching is a craft practiced by a group of professionals and as such there are certain moral dictates which influence, or ought to influence, techniques employed by individual instructors when teaching students. These moral dictates are evident to any morally conscientious practitioner and violations of the dictates are also similarly recognized. People who are not members of the profession cannot be expected to be aware of the moral dictates which bind members of a particular discipline. Consequently, students, politicians, journalists and other non--academics cannot be expected to recognize those occasions in which serious transgressions of the profession's moral commitments have occurred. It may be that too many academics are in fact unaware of the moral responsibilities of university instructors and this is certainly virgin territory for empirical research. In any case it is again unlikely that students can make informed judgments about whether or not an instructor has effectively practiced the craft of teaching in a morally responsible way. In all but the most obvious cases of professional irresponsibility (such as the acceptance of a bribe), the latitude guaranteed to the instructor through the auspices of academic freedom is known only to fellow members of the professional community. If an instructor has failed to keep abreast of his discipline it is not the novice who will first recognize that fact but the experienced professional who recognizes his colleague's ambivolance

toward new developments in the field and general disinterest in keeping his students informed on state of the art material.

In light of the preceeding discussion the following sources of relevant evidence become evident:

1. Student examinations: Examinations display levels of student competency. Such information is relevant to determining the amount of academically respectable material presented to the students. The grade awarded to the student is also an indicator of what the instructor judged to be significant levels of achievement in his or her course. This can be compared to how experts in the field would evaluate the same level of achievement for a similar course.

2. Discipline--specific specialist observation: Only historians or perhaps historically--minded sociologists, political scientists and humanists are sufficiently versed in the subject matter of say, history, to recognize if the instructor's presentation is likely to increase student understanding of certain given material and if such material is academically significant. Only knowledgeable teaching fellows are in a theoretically sound position to make a rough estimate of the teaching effectiveness of a colleague.

3. Course syllabus: Course syllabii illustrate the sort of material the instructor claims are relevant at this level of student preparation. Colleagues can determine whether such material merits continued study.

4. Neurphysiological change: When the scientific study of humans advances to the point that we can identify neurphysiological change as a consequence of subtle pedagogical input this would no doubt be the most exciting area of study of all.

I make no claims about the exhaustiveness of this list. There may be other sources of evidence--particularly certain subsets of item four above. However at this time we have no theoretical justification and no reason to identify any other class of observations as relevant to the assessment of effective university teaching. Clearly, all student evaluations of teaching effectiveness have been shown to be wholly without any theoretical support. Even general expressions of student satisfaction have never been shown to be conceptually linked with teaching effectiveness. If an evaluation of teaching effectiveness is to occur in a fair and maximally objective way, then evaluators must limit themselves to working only with those sources of evidence for which there exists clear and uncontroversial support.

References

1. See for example, Norwood Hansen, Patterns of Discovery, (New York: Cambridge University Press, 1958); See also R. L. Gregory, Mind and Science (New York: Cambridge University Press, 1981) Chapter 1; Paul A. Wagner, "Review of R. L. Gregory's Mind and Science," Cognition and Brain Theory, Vol. V No. 3 Summer, 1982. pp. 294-299.

2. See Norwood Hansen, Patterns of Discovery, see also, Peter Achinstein, Concepts of Science (Baltimore: John Hopkins University Press, 1968) especially chapters 5, 6; Stephen Toulmin, Human Understanding (Princeton: Princeton University Press, 1972); H. I. Brown, Perception, Theory and Committment, (Chicago, University of Chicago Press, 1980); Bas van Fraassen, The Scientific Image, (New York: Oxford University Press, 1981); Clark Glymour, Theory and Evidence, (Princeton: Princeton University Press, 1980).

3. See for example, Thomas Kuhn, The Structure of Scientific Revolutions, (Chicago, University of Chicago Press, 1962); I. Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in Criticism and the Growth of Knowledge, I. Lakatos and A. Musgrove (ed.), (New York: Cambridge University Press, 1970) pp. 91-196; J. H. Fetzer, Scientific Knowledge (Boston: D. Reidel Publishing Co., 1981).

4. While Scrivens made such an observation in an evaluation workshop it would be unfair to leave the reader with the impression that Scrivens is wholly insensitive to the ethical and content aspects of teaching. He is not. In fact in his article entitled, "Summative Evaluation," in Jasen Millman (ed.) Handbook of Teacher Evaluation, (Beverly Hills: Sage Publications, 1981).