DOCUMENT RESUME

ED 031 820 24 EA 002 551

By-Trow, Martin
Methodological Problems in the Evaluation of Innovation.
California Univ., Los Angeles. Center for the Study of Evaluation of Instructional Programs.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Report No-CSE-R-31
Bureau No-BR-6-1646
Pub Date May 69
Contract-OEC-4-6-061646-1909
Note-32p.; From the Proceedings of the Symposium on Problems in the Evaluation of Instruction (Los Angeles, December, 1967).
EDRS Price MF-$0.25 HC-$0.70
Descriptors-*Behavioral Science Research, College Curriculum, *Educational Innovation, Educational Objectives, Educational Policy, Educational Research, *Evaluation Methods, *Experimental Curriculum, General Education, *Higher Education, Organizational Climate, Political Influences, Research Problems, Vocational Education
Identifiers-Hawthorne Effect

This report outlines three aspects of research and innovation in higher education. First, the relevance of social science research methods to educational policy is discussed in relation to the bureaucratic and democratic distribution of power in higher education. Second, research and innovation are discussed and applied toward determining the proximate and ultimate goals of an educational institution. Finally, the influence of experimental research on educational processes is evaluated. Some research methods to explore these difficulties are outlined. Suggestions are made to make innovative research more enlightening by emphasizing its use in clarifying and identifying processes and procedures in education. Related documents are EA 002 410 and EA 002 536. (MLF)

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds.

METHODOLOGICAL PROBLEMS IN THE EVALUATION OF INNOVATION

Martin Trow

University of California, Berkeley

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSE Report No. 31, May 1969
University of California, Los Angeles

METHODOLOGICAL PROBLEMS IN THE EVALUATION OF INNOVATION

Martin Trow

Sources of Innovation in Higher Education

Currently there is considerable ferment in American higher education arising from widespread discontent with present arrangements and practices. This dissatisfaction has its roots in a set of developments in higher education and in the larger society which are changing the character and functions of our colleges and universities and at the same time changing expectations of what they should be doing.

These forces affect individual institutions in very different--or even opposite--ways. For example, the steady increase in the percentage of the population which attends college raises the academic quality of entrants to the more selective institutions, while it brings to less selective institutions large numbers of students who are there, at least initially, because there is not much else for them to do. The presence of large numbers of relatively unmotivated students in colleges which have no strong academic traditions poses problems similar to those which gave rise to the transformation in curriculum and in teacher-student relations in our high schools earlier this century. They are the problems, in brief, of generating in the classroom the interest and motivation which one could no longer assume the student brought with him. The concern for relating the curriculum to the lives and interest of the students, rather than to

a traditional body of knowledge or the specialized interests of the academic disciplines, underlies many of the current efforts to change the form and content of instruction, especially at the undergraduate and introductory levels.

There are other forces which are making our traditional forms of education less and less satisfying, in the graduate and professional schools as well as in the undergraduate liberal arts colleges. The rapid growth of knowledge makes the traditional syllabus obsolete, while simultaneously weakening the traditional boundaries of the academic disciplines. Closely related are changes in professional education as an increasingly wide range of knowledge becomes directly relevant to effective professional practice; the growing role of the social sciences in the education of physicians, lawyers, engineers, architects, and city planners is a case in point.

Whatever its sources, the ferment has led to a variety of new approaches to higher education. These include the sweeping innovations in the organizational forms of higher education, such as the consortium of institutions in California at Claremont or in New England in the Connecticut valley; the single institutions which embody some distinctive organizational principle, such as Santa Cruz's collegiate structure; the varied means of approaching what used to be called "general education;" and the latest effort any one of us may make to create a new course around a problem, or a cluster of

disciplines, a new way of using teaching assistants, or the new technology of electronic instruction.

There are many kinds and degrees of innovation, and the problems of assessment of these varied efforts obviously differ. Though it is difficult, I will try to say something about educational innovations, regardless of how far-reaching in intent they are or of where initiative lies. My emphasis will be less on the technical problems of evaluative research--the relative strengths of different modes of investigation or different strategies of analysis--than on the characteristics of the phenomenon being studied and assessed and on the social context in which they are embedded. What forces give rise to an innovation? What are the criteria of its success? Who cares whether or how it is assessed? These are problems for the researcher which often override the knotty difficulties of how to measure change or the influence of a clique of friends.

I would like to address myself to innovations in the curriculum and in the modes of teaching and learning, rather than to innovations in broader organizational forms, which I think involve a somewhat different set of "methodological problems."

Some Functions of Educational Innovations

Innovations in instruction in higher education arise most often out of some felt sense of the inadequacy of existing arrangements,

and very often from sheer boredom with what one has been doing. We are always tinkering with our courses or with the curriculum, even when they are working reasonably well. And while proposals, whether for a new college or for a new course, are usually justified as promising some improvement on what is being done, very often we know or strongly suspect that what is proposed recommends itself not so much on its promise of betterment as on the certainty of its being different.

An innovation is a break with routine and habit; it disrupts unreflective ways of thinking, feeling, and behaving; it requires a heightened measure of attention and interest in the matters at hand; it forces the participants, and especially the creators, to think in fresh ways about familiar subjects, to reconsider old assumptions. Above all innovation dispels, if only briefly, the fog of boredom that hovers over everything we do in our classrooms. Habit and routine are extremely useful in allowing us to do a great many necessary things without having to think about them, while freeing our minds and energies for other more demanding matters. But when habit and routines begin to encrust educational structures and processes, life, thought, interest, and creative imagination go out of them and they become boring to us and to our students. I think we know intuitively that boredom is a greater enemy of education than ignorance or error or even stupidity, that as an enemy it is rivaled only by

dogmatic authority. And if boredom is a chief enemy, innovations and change are our chief weapon against it because they can break through routines and release fresh energy, imagination, and inquiry.

I am suggesting that innovations in education justify themselves by their intrinsic qualities almost without regard to their outcomes. Indeed, innovation goes on constantly. For the most part it does not advertise itself by name, often because the innovator does not need additional resources and because he does not want to become entangled in the cumbersome machinery through which formal changes in the curriculum are made.

Whether advertised or not, it is important that innovation is commonly done for its own sake and only secondarily for its outcomes. Because that fact greatly reduces the relevance of systematic evaluation of innovation, it reduces the significance of the manifest functions of evaluation--to tell the innovator what he has achieved and how successfully--as compared with its chief latent function--to legitimize an innovation and contribute to its continuation and extension. Innovations will be made--with or without evaluations--almost regardless of their nature simply because we enjoy making them. From this perspective, evaluation studies are aimed less at the innovator than at funding agencies or course committees which can support or limit the life or scope of the innovation. Such studies thus are typically directed at expensive innovations or those

which have a broader impact on other parts of an institution and thus involve persuading others that the innovation has value and should be supported. Innovations that are inexpensive or are confined within one department or one course usually are not evaluated; they are just enacted.

To emphasize the latent functions of some social pattern or practice like evaluation is implicitly to minimize the significance of its manifest functions. A large part of evaluation in education is best understood as a form of persuasion directed at powerful people who make decisions and control resources. But this is not to say that evaluation studies need be nothing more than devices for legitimation and persuasion. In modest ways evaluation can help shed light on educational practice, and perhaps help us see what an innovation actually consists of as well as what it achieves. But the context and function of such studies affect the way we conduct the studies and how much confidence we can place in the findings, and thus are deeply implicated in methodology.

I would therefore like to discuss three aspects of research and innovation in higher education. First, the political context and significance of evaluation. Second, the educational and research problems posed by the diffuseness of the intended outcomes of education, including its innovative forms and the long delay beyond the college years before many of these outcomes manifest themselves.

Third, the great difficulty, especially in innovative courses, of distinguishing the special circumstances surrounding their creation and adoption from the other characteristics of the innovation which may recommend it to others and to its institutionalization.

The Political Context of Research in Higher Education

We see in American higher education a growing sense of the relevance of systematic research procedures along with a considerable hostility to social research and suspicion of the educational implications of its findings. Paradoxically, both the growing need for such research and the wariness of it rise from similar sources. The rapid growth and democratization of higher education--a growth which brings into our colleges an enormous number and variety of students whose values, motives, and purposes are strange to the academic man--leads to the extension of social research in many colleges and universities. Moreover, conditions in the large public colleges and universities make it difficult to establish the old personal relation of student and teacher, and thus for the faculty member to know his anonymous students in any real sense.

Increasingly, and often for much the same kind of practical reasons which prompted the social surveys of the 19th Century, educators are turning to social science to learn the facts about their students that are no longer directly knowable by the teacher or administrator. But this return is met with the same ambivalence among

cultivated men as was the earlier development of social research. It threatens the academic man's role as an intellectual and as an interpreter of his own social experience; it asserts that much that is of importance--not only in the wider society but in his own classroom and in the students' residence halls--no longer can be adequately known and understood by the man of intelligence and sensibility. The suggestion--often made tactlessly by social scientists--that the professor of humanities cannot grasp the social processes going on around him without the aid of the social scientist's special skills and techniques is frequently met with hostility and resentment. The very existence of social research on campus, as some professors have candidly stated, is an insult to their intelligence. Their response, made perhaps with more feeling than logical consistency, is to doubt that social science is more than a pretentious fraud and, at the same time, to fear its manipulative consequences.

But social research threatens not only the intellectual competence of academic men regarding their teaching functions; it is also felt by some to be a threat to liberal education. Colleges where educational practices and arrangements are seen as embodying the values of the institution, instead of merely facilitating their attainment, are likely to be inhospitable to the notion of applying the findings of social research. To the extent that its practices

have become highly institutionalized, charged with value in themselves, a college will resist conscious planning based on rationalized procedures and data. Such an institution is likely to rely on committee deliberations as more likely to preserve the primacy of the substantial values. By contrast, a college committed to the achievement of easily measurable goals, and which is prepared to measure and modify its practices against the criterion of the efficient achievement of these goals, is more likely to sponsor and apply social research against whose findings elements of the organization can be evaluated.

Liberal education is a substantial value in itself. It <u>is</u> the practices and relationships and patterns of behavior that enter into it, much more than it is some nebulous "outcome," difficult if not impossible to measure. By contrast, vocational and professional education is to a much greater extent instrumental and goal-oriented--the outcomes are measurable in skills and knowledge acquired, examinations passed, diplomas earned. The colleges and the parts of large universities that are deeply committed to liberal education have been less likely to welcome or apply social research which touches on their core values and activities than have those organizations or parts of organizations whose practices are instrumental to some more clearly defined or measurable goals.

Typically, in American colleges and universities, power is distributed in extremely complicated and obscure ways among the administrators, the faculty, the trustees, and various important constituents,

such as alumni, the current body of parents, and--in the case of public institutions--the legislature or other sources of public funds. The question of what is manipulable and by whom is itself highly uncertain, at least as difficult to know as patterns of student behavior, which may be the nominal subject of investigation. Every organization is to some extent a polity, in which political processes determine who can initiate what events, who can veto them, and whose consent must be gained before policies are put into effect or sabotaged. Some studies of internal organizational processes have been done within formally bureaucratic organizations (e.g., business firms) and within formally democratic organizations (e.g., trade unions and political parties), but almost nothing has been done by way of studying the political processes within institutions of higher education. These institutions are in part bureaucratic and in part democratic, combining the principles of hierarchy and colleagueship in varying degrees. I am not suggesting a design for the study of colleges and universities as political structures, but rather that the relevance of social science to educational policy cannot be discussed without recognizing that policy recommendations within colleges are quickly transformed into political issues.

A highly rationalistic conception of the relation of research to policy obscures the political character of a college and of recommendations to it: those who hold such conceptions are continually

surprised and indignant when the institution does not take the "reasonable" course of action suggested by research. A director of the Bureau of Institutional Research at a large mid-western university has described, with becoming candor, actions taken by faculty committees in two cases in which his bureau conducted research on the issue in question--both actions were at variance with the apparent indications of the research. He observed, with more sadness than anger, that "actions such as these represent one of the frustrations of a person in institutional research. The mere establishment of an institutional research unit does not in itself guarantee that decisions will be made on a more realistic, objective, and reasonable basis. As you can see, even in our institution with its long tradition of faculty-oriented institutional research, faculties and faculty committees have been known to make decisions on other than a purely objective basis."

Without describing these cases in detail, I can report only that the research center's recommendations are "realistic and objective" on the basis of a rather narrow conception of educational efficiency, and that faculty members with other values regarding education might well see such a research report as a political document and oppose it as such. But the claim to objectivity denies the value implications of the research and makes opposition to it mere pigheadedness, or in the words of this research man, "stubborn resistance to change."

This in itself tends to excite suspicion of all social research among faculty members whose values are frequently at variance with those implicit in, but denied by, offices and bureaus of research--a point to which I would like to return a little later.

The general principle that policy recommendations (whether they are or are not based on social research) are in most cases immediately transformed into political issues alerts us to a number of politically relevant factors intervening between research and implementation. The distribution of power in colleges and universities is more diffuse than in most formal organizations. The principle of bureaucracy tends to centralize formal power and authority at the top of the hierarchy, while the principle of colleagueship tends to spread it more widely among the faculty. There is some evidence of the existence of a long-range trend toward the diffusion of power by means of a strengthening of the principle of colleagueship and of faculty participation in the government of colleges and universities. The AAUP, for example, finds that over the past several decades faculty influence in most of the colleges they have been studying has been growing. There seems little doubt that this tendency is a result of the strong efforts American colleges and universities are making to upgrade themselves to the level of the more distinguished colleges and universities where the principle of colleagueship is most strongly and influentially established. One result of this tendency is for the interests and values of the faculty

to become more widely and more directly involved in the application of social research to educational policy. This in turn makes it increasingly difficult for administrators to act with authority, even on the basis of research findings and recommendations.

The interests of the faculty are touched at many points by proposals to modify the structure or content of an educational program. Clearly, areas of investigation vary in the degree to which they visibly impinge upon the interests of the parties concerned. In general colleges will be more receptive to applied research on issues further removed from the interests of those who make the decisions--more hospitable, for example, to research on student life than on faculty authority, to research on the social implications of residence hall architecture than to studies of the distribution of power in college and university departments.

In the United States the bulk of applied research in higher education has been carried out by fact-finding agencies within the colleges and universities--by assistants to the president, by deans or assistant deans, by testing offices, and increasingly by offices of institutional research. The line between social statistics and social science is a fine one and lies in the shift of a passage in a report. Absence levels, for example, may be indicators of underlying social and institutional processes; the next step is to study these processes more directly. If this step is taken relatively

rarely, it is partly because the people who do this kind of research for colleges and universities rarely are social scientists or have an interest in organizational analysis, and partly because of the suspicion with which research is viewed by important parts of the faculties of many institutions.

Some of this suspicion has a different basis from that which arises from the dispute between humanists and social scientists over the relative power of science and sensibility for interpreting social life, but its effects are similar and reinforcing. The suspicion arises from a profound struggle that goes on within many institutions and takes many different forms, a struggle between those committed to some ideal of liberal education--to the development of the intellectual powers of the individual, of his breadth of vision, independence of mind, and critical faculties--and those primarily interested in education for extrinsic ends, for social and vocational skills. The suspicion of research held by many humanists is that in this struggle, basically a political struggle over the means and ends of education, research is usually on the side of the vocationalists.

It is thought to be so, not only by virtue of the kinds of people who do it but also by virtue of the very kinds of data they collect. For while the indicators of success of a liberal education are likely to be vague, difficult if not impossible to measure, and scarcely distinguishable from the effects of all the other experiences a

student has had in his life, the indicators of successful training are the kind of performances that testing offices and offices of institutional research can measure. ~~There~~ Their cognition of this by those faculy members committed to liberal education, and the suspicion that arises from it, partly explains the mechanisms that surround offices of research to insulate them from the core values and activities of the faculty; for example, their subordinate status and their definition by the institution as technical agencies gathering statistical information primarily for administrative uses, rather than for basic research into the nature and processes of higher education.

The criteria and indicators of "success" of educational practices or innovations that are employed in educational research are elements in the academic-political controversies on many campuses. They affect the forms that research takes and the reception it gets--that is what happens to it.

Proximate and Ultimate Goals of Education

Some "outcomes" of education are easily measured, and for that reason, as well as others, they are commonly measured. Among these are the student's grade point average, drop-out or transfer rates, achievement of graduate scholarships and higher degrees. These matters are part of almost every research into higher education,

not only because they lend themselves to easy and systematic measurement, but also because they are important in themselves. Grades are not merely an "index" (however weak) of what has been learned; they are also an important determinant of the individual's future opportunities and life chances, among them his chances of gaining admission to a good graduate school. Acceptance by a good graduate school is an even more important determinant of a man's chances of making significant contributions to science or scholarship.

But whatever their objective importance, which is very great, grades and higher degrees are inadequate measures of the outcomes of educational experience for many reasons. They do not measure the whole of what some men wish education to do to or for students. They are poor measures, for example, of the success of a liberal education in refining sensibilities, developing capacities for critical and independent thought, or the use of reason and evidence in everyday life, or the enhancement of the individual's capacities for enjoying life and making fruitful contributions to it. Some men want these great goods to flow from a scientific and technical education as well. The difficulties in discovering whether indeed an education has these effects are several.

In large part, these qualities of mind and spirit do not show themselves, during the college years, but may be laid down then as potentialities which bear fruit in later life and career. They are,

for the most part, exceedingly difficult to measure systematically, however much we pride ourselves on our ability to recognize their presence or absence in others. Moreover, these qualities are not only valued outcomes of formal education, but also the products of the whole of man's genetic equipment and life experience. Even if we could measure them with some precision and confidence, how are we to distinguish the part played by formal higher education from all the other more enduring and emotionally weightier influences on a man's life and character?

In a word, then, the most important and truly valued outcomes of higher education are extremely difficult, if not impossible, to assess. As a result, many institutions, usually those with the least firm educational purposes and the least distinctive character, fall back in their self-assessments on those presumed outcomes of higher education that are most easily measurable. And, in a familiar translation of necessity into virtue, such an institution may define its aims in terms of what can be measured, and to shape and justify its practice in terms of its success in reducing the drop-out rate, increasing the number of fellowships its graduates earn, and the like.

What are the alternatives for the institution that does not want to reduce its educational aims to the level of the most easily measured of student characteristics? Matters are not quite so hopeless

as my remarks above may seem to suggest. There are things that research can do to help an institution assess its success in achieving its most profound and not merely its most proximate aims. For example:

1. We are not confined to the study of the most obvious and easily measurable outcomes of education. There are ways to explore changes in basic values and attitudes of students, and even aspects of their personalities which education aims to modify over the college years; to explore changes in life plans and the conditions and experiences in the institution which give rise to them; to attempt at least to study such subtle matters as creativity and independence of mind and judgment.

2. These are all to a considerable degree a product of the student's life experience before coming to the institution. To some degree we can assess the extent to which they are already present at entrance, so that we can make some assessment of the relative efficacy of different educational practices during the college years in developing (or inhibiting) these qualities.

3. We can do far more than has been done to follow our graduates into their adult careers to see what happens to them there, and to see if we can make even tentative inferences about connections between their adult careers and their college experience.

The Influence of Experimental Research on Educational Practice and Its Outcomes

Different forms of social investigation vary in the extent to which they affect the educational processes that they aim to illuminate. A survey of a college's alumni presumably would have little direct influence on the faculty and students at the institute at the present time. Questionnaires distributed to entering freshmen probably will have relatively little effect on their subsequent behavior, though repeated questioning about a given issue--say, the question of student-faculty relations--might be expected to increase the salience of that issue in the minds of the students. But experimental research-linked changes in the curriculum are likely to have very marked consequences for the teaching-learning process over and above those effects which the alterations are specifically intended to achieve. It may be worthwhile to consider for a moment the problems such experiments pose for research designed to assess their effects and effectiveness.

First, there are the difficulties, already discussed, of measuring the genuinely desired outcomes and of disentangling them from the manifold extraneous influences of life and time outside the experimental classroom. Experiments share this difficulty, as I have suggested, with other forms of research into education.

In addition, experiments in education, like social experiments in general, pose special difficulties for research, in that they introduce into social situations powerful forces over and above those purposefully introduced by the experiment. These "other forces" affect the outcomes of the experiment in ways that are very difficult to separate from the effects of the "intended" experimental procedures, that they so closely resemble them. The general phenomenon to which I am referring has become known as the "Hawthorne effect," after the famous experiment on worker productivity at the Hawthorne plant of the Western Electric Company in the late 1920's. That study showed that the experimental situation itself, independent of the purposeful manipulation of the situation, modified social relations, group morale, and individual motivations among the subjects in ways that affected their performance, in most cases for the better.

This phenomenon has become widely associated with the independent and common observation that in education no experiments fail, so that it has been seriously suggested that one educational strategy would be to "institutionalize the Hawthorne effect" by making "experimental" innovations a regular part of school or college administration. This advice has not been widely adopted because institutions are made as unhappy as individuals are by a steady diet of innovation; it puts a strain on lines of communication and authority, makes the coordination of the different parts of the institution more difficult, and

makes life less predictable and thus more unsettling and anxiety-arousing for the individual. The gains of educational innovations may be worth all this, but before recommending such a strategy which dissolves the distinction between "action" and "research" by making the research itself the action, it may be worth considering what are the forces involved in such "experiments" to see if indeed they can be made part of the institution's regular procedures without their unsettling side effects. Put another way, what are the sources of their evident power to raise performance?

1. One of the forces generated by a classroom experiment is to make the "subject" students feel somehow distinctive, a "special" group getting special attention. This effect of the experimental situation was noted at Hawthorne, where it presumably generated among workers there the special group morale and commitment to the task that resulted in their higher individual performances.

2. Quite distinct from that process, however, is the fact that experimental courses are customarily instituted and taught by imaginative teachers, who have given an extra measure of thought and effort to the pedagogical problems they face--the innovators themselves, one may guess, are probably better than average teachers. This cannot help but play a part, perhaps the major part, in their customary "success."

3. Not only is the self-selected staff of an experimental class likely to be more gifted than the average but they are also likely to have a strong interest in the success of their "experiment," and to communicate that interest through the enthusiasm with which they tackle the course. Enthusiasm for a subject is a well-known characteristic of the successful teacher, even in more routine courses. Coupled with the innovative character of an "experiment," it is a powerful pedagogical force.

4. Typically, if not uniformly, "experimental" courses have been assigned larger amounts of the institution's resources than have comparable "routine" courses. The ratio of teachers to students is higher, and the amount and intensity of student-teacher interaction is commonly greater in "experimental" than in routine courses. This also helps educational "experiments" to succeed, both through the more thorough way in which the course material can be covered with each student and through the higher levels of student motivation that teacher attention can generate.

Much of the success of an "experimental" course is related to the fact that it is a break in routine which forces a higher level of imagination and energy from the staff and excites it in the student. The sheer innovative character of such an "experiment," coupled with its typically rich endowment of resources by the institution, almost ensures its success independent of its purposeful content.

But the problem for research which aims at assessing the worth of an educational innovation is clear: how to distinguish the experimental effects from the designed or purposeful effects. It may be argued that the time to assess an innovation is when it is no longer an innovation, when it has become routinized and no longer can call forth the special energies, resources and enthusiasms of an "experiment." The trouble is that an institution usually wants an assessment of an experiment in the curriculum before it has committed major resources, before it has made the necessary organizational adjustments, and before it has persuaded people who did not initiate it to staff it.

I have emphasized the difficulties for research in assessing the worth of a curriculum experiment, but I do not want to exaggerate them. Research methods of several kinds can be employed to explore the workings and outcomes of an experimental course, and such research may be of real value to the institution so long as the policy-makers recognize the special characteristics of educational experiments that make them so difficult to assess. For one thing, the degree of "success" of such a course, whatever its sources, can be tested at its conclusion by using the ordinary indicators of comparative performance on examinations, or more subtle indicators of intellectual powers and creativity that might be devised. Another approach is to try to identify the pedagogical forces set loose by an innovation by

subjecting the experimental course to close and continuous observation, aiming to see what elements in it call forth ghe greater motivation and effort that I assume will be observed. Such observation, of course, should also be accompanied by parallel observation in "ordinary" classes covering the same or comparable materials, to allow something approaching a comparative analysis of the observational data. It may well be that such observation will allow the research to identify aspects of the course--pedagogical devices, organization of the subject, or whatever--which, though not explicitly "intended" by the innovators, appear to be particularly successful, and which might be more widely introduced into the curriculum on a regular basis. In a sense this would be an effort to separate the pedagogical forces associated with innovation from innovation itself. It would be an attempt not to institutionalize innovation but rather to identify those of its elements which are not dependent on the presence of the innovator or extra resources. Knowledge of the genuinely effective aspects of educational practice might liberate institutions from reliance on the specific educational forms in which they manifest themselves, allowing the invention of new forms which embody the effective processes in more effective or less expensive ways. To my knowledge, this kind of observation has not often been done on a systematic basis in educational institutions, and while the value of such observations is heavily conditioned by the skill and sensitivity of the observer, it very much warrants trial.

Illuminative versus Evaluative Research

I have been discussing thus far some of the problems of evaluative research in higher education: difficulties rooted in the suspicions of humanists and the conflicts within faculties; difficulties in the criteria we use to assess educational efforts and in the remoteness of ultimate goals from proximate outcomes; difficulties in disentangling the unique qualities of innovative teaching procedures from their enduring and transferable qualities. But I feel an obligation here to end on a more hopeful note: to suggest that these difficulties are superable, that they are worth the effort needed to deal with them.

The first issue has to do with the institutional context of evaluation: who does the job, to whom does he report his findings, and what is done with his report? Insofar as evaluation is done by a research arm of the administrator, who reports to the administration regarding the value of certain aspects of the curriculum, the research enterprise is likely to face considerable suspicion and hostility from the faculty. As I have suggested, much of the suspicion is merited, since evaluation must be predicated on educational values, however disguised as science, and these values are very often--I might almost say chronically--in dispute. The way for evaluative research to meet this suspicion involves two changes in the character of such research. First, research on innovative efforts must be seen from

the beginning as "illuminative" rather than as "evaluative" in the narrow sense. It must recognize that the value of innovation also comprises the rewards gained by the faculty members who create it and are not confined to its easily measured outcomes. Second, researchers must recognize that these outcomes bear only a remote relation to the ultimate impact the faculty member may be hoping to have on the minds, characters, and lives of his students. This means that such researchers must forego the dubious pleasure of awarding gold stars and demerits to academic innovators, but must try instead to serve them. Research on innovation can be enlightening to the innovator and to the whole academic community to whom research reports are properly addressed by clarifying the processes of education and by helping the innovator and interested other parties to identify those procedures, those elements in the educational effort, which seem to have had desirable results. Such research may involve a comparison of proximate results, such as examinations, papers, and so forth, with those produced by other more conventional courses. It may also involve close semi-participant observation of the course in an effort to identify the operative social and psychological mechanisms which the innovative procedures create (often beyond anyone's intention) which engage the interests and efforts of students and open them to the instructor's attempts to transmit skills, broaden horizons, or deepen understanding. Precise techniques

of inquiry are not at issue here; we know pretty well their characteristic strengths and limitations. What is important is that the research be seen to be in the service of the innovative enterprise, and not sitting in judgment on it. And for that, it must accept its own tentativeness and function as a facility of the faculty, not as a part of the administrative apparatus. The formal status of the researcher or the research group, who employs him, to whom he addresses his findings, and how he avoids being drawn into academic controversies, are crucial here, though circumstances differ enough that no set of recommendations on these matters can possible apply to all institutions.

In considering the gulf between proximate indicators of the results of educational innovation and their long range goals, wisdom resides, I believe, in a decent regard for the limits of research. What are the qualities that make creative engineers, resourceful businessmen, thoughtful and responsible citizens, men of independent mind, moral sensitivity, and aesthetic sensibility? What relation does college performance bear to these qualities? And what influence do specific educational arrangements have on what men do and what they are in their lives? A consideration of the kinds of men who have been exposed to the most varied kinds of higher education, or to none at all, should make us pause before we give any ready answer.

When we return, as does the teacher himself, to the student before him, we may attend to what we see not merely as a most imperfect

indicator of future achievement of qualities, but as of intrinsic importance. On one hand it is important that students be able to learn and be able to demonstrate that they have learned assigned material--important for its effect on the range of possibilities that are open to the successful students but closed to the academic failures. On the other hand it is at least as important whether students are bored or engaged, committing their energies or cooly withholding them, fulfilling obligations or freely involving themselves in learning. And these things as we know can be affected by educational arrangements and procedures--however constrained by deeply set qualities of mind and character that the student brings with him to college and that remain with him unaffected there. We also know with Woodrow Wilson that:

> The real intellectual life of a body of undergraduates, if there be any, manifests itself, not in the classroom, but in what they do and talk of and set before themselves as their favorite objects between classes and lectures. You will see the true life of a college--where youths get together and let themselves go upon their favorite themes--in the effect their studies have upon them when no compulsion of any kind is on them and they are not thinking to be called to a reckoning of what they know.

We know also that the life of the student outside of class can be influenced by our efforts. The innovator can see some of this, but he is busy teaching. The researcher can see more, much more. He is trained to see just those things, and he is less constrained to see

that he hopes to see. The illumination of educational innovation through systematic research can be in large part the identification of those educational processes that can be linked to the innovation--the processes of learning and growth that go on inside and outside the innovative classroom, laboratory, or residence hall.

Finally, with regard to the uniqueness of innovation and the special resources of talent and imagination frequently available, it may be that research should attend precisely to those qualities of abundance, rather than trying to "partial them out" in assessing the effects. It may be that what we should aim for is not so much the routinization and institutionalization of successful experiments but a climate and organizational arrangements which make innovation easy and frequent. If, as I suggested at the outset, innovations recommend themselves for their intrinsic qualities rather than for their putative outcomes, if they are our chief weapon against boredom and routine, then the real research effort should be directed toward the effort to "evaluate" them once in being. This posture is completely compatible with the aim of illuminating their processes and proximate gains. We can want to encourage innovation, while recognizing that some experimental efforts will be more successful than others, judged by their own and by broadly accepted criteria. We need not set aside all "academic standards" or notions of craftsmanship and achievement in a wholly unreflective celebration of

academic spontaneity. There is enough anti-intellectualism afloat today, both inside and outside the academy, without social research contributing any more. But here we come very near to what is perhaps the central problem for the student of educational innovation. For in innovation we are very often dealing with "enthusiasm" on the part of innovators and sometimes of their students as well. On one hand, this enthusiasm means heightened attention, alertness, involvement, commitment, creativity; on the other, the danger of enthusiasm lies in the passion of the true believer and of his terrible certainties. If our studies of educational innovations can illuminate those forces which are respectively the chief instruments and enemies of education, we can perform a very considerable service to our students and to our innovative colleagues, and to the enterprise of learning.