R E P O R T   R E S U M E S

ED 015 534 EA 000 943

DEVELOPMENT, DIFFUSION, AND EVALUATION.

BY- GUBA, EGON G.

NATIONAL INST. FOR THE STUDY OF EDUCATIONAL CHANGE

PUB DATE OCT 67

EDRS PRICE MF-$0.25 HC-$1.80 43P.

DESCRIPTORS- *DEVELOPMENT, INFORMATION SYSTEMS, *DIFFUSION, EVALUATION, CHANGE AGENTS, *EDUCATIONAL PRACTICE, *RESEARCH UTILIZATION, *EDUCATIONAL RESEARCH, RESEARCHERS, RESEARCH AND DEVELOPMENT CENTERS, REGIONAL LABORATORIES, EDUCATIONAL INNOVATION EVALUATION TECHNIQUES, PORTLAND, TITLE III PROJECTS,

THE KNOWLEDGE GAP BETWEEN INITIAL RESEARCH AND FINAL USE IS DISCUSSED IN TERMS OF THE FOUR STATES OF THE THEORY-PRACTICE CONTINUUM (RESEARCH, DEVELOPMENT, DIFFUSION, AND ADOPTION). THE TWO MIDDLE STAGES ARE EMPHASIZED. RESEARCH AND DEVELOPMENT CENTERS, REGIONAL EDUCATIONAL LABORATORIES, AND TITLE III PROJECTS ARE SUGGESTED AS AGENCIES RESPONSIBLE FOR THE DEVELOPMENT FUNCTION. FAILURE OF THE DIFFUSION FUNCTION TO OPERATE SATISFACTORILY IS ATTRIBUTED TO THE LACK OF AN ACCEPTABLE STRATEGY. SUCH A STRATEGY WOULD CONTAIN THESE ELEMENTS--(1) ASSUMPTIONS CONCERNING THE NATURE OF THE PRACTITIONER WHO WILL BE EXPOSED TO THE STRATEGY, (2) ASSUMPTIONS CONCERNING THE END STATE IN WHICH ONE WISHES TO LEAVE THE PRACTITIONER, (3) ASSUMPTIONS ABOUT THE NATURE OF THE AGENCY OR MECHANISM CARRYING OUT THE DIFFUSION ACTIVITY, AND (4) ASSUMPTIONS CONCERNING THE SUBSTANCE OF THE INVENTION. THE CONCEPT OF EVALUATION IS CHANGING RAPIDLY, AND THE METHODOLOGIES CURRENTLY USED FOR EVALUATION NEED REPLACEMENT. TRADITIONAL EVALUATION HAS FOUR CHARACTERISTICS WHICH ACCOUNT FOR ITS LIMITED UTILITY--(1) EVALUATION DATA ARE USUALLY AVAILABLE ONLY UPON THE TERMINATION OF THE EVALUATIVE PERIOD, (2) EVALUATIVE DATA TYPICALLY AFFORD ONLY A RETROSPECTIVE VIEW, (3) THE ASSUMPTIONS ON WHICH EVALUATIVE DESIGNS ARE BASED IMPOSE A SERIES OF CONSTRAINTS ON THE EVALUATION, AND (4) THE CONSTRAINTS IMPOSED CREATE A LABORATORY CONDITION WITHIN WHICH THE TREATMENT IS TESTED. THIS PAPER WAS PREPARED FOR THE UCEA CAREER DEVELOPMENT SEMINAR (17TH), COSPONSORED BY THE UNIVERSITY COUNCIL FOR EDUCATIONAL ADMINISTRATION AND THE UNIVERSITY OF OREGON (PORTLAND, OREGON, OCTOBER 22-25, 1967). (HW)

DEVELOPMENT, DIFFUSION,

AND EVALUATION

Egon G. Guba

A Paper Prepared for the

Conference on Knowledge Production and Utilization in Educational Administration: Role Emergence and Reorganization

Co-sponsored by the University of Oregon and the University Council for Educational Administration, Portland, Oregon, October 22-25, 1967

THE NATIONAL INSTITUTE FOR THE STUDY OF EDUCATIONAL CHANGE
825 East Eighth Street -- Bloomington, Indiana

More than a decade ago I was a self-styled "expert" in the area of administrative staff relationships. My colleague at the University of Chicago, Jack Getzels, and I strove mightily to put the terms "nomothetic" and "idiographic" into the vocabulary of every practicing administrator in the country. I recall that we made a lot of speeches on the subject, Jack and I, and usually there was a question or discussion period following. Almost inevitably this comment would come from someone in the audience, "What you say seems to make some sense, although I'm not sure I really know what you're talking about. Why don't you fellows come down out of your ivy tower and tell us about your ideas in language that we can understand? How about showing us how to apply those ideas 'on the firing line'?"

"Well we would say, 'Practice is hardly our concern. We don't know what the practical problems are. It's up to you administrators who have to deal with these problems every day to make the application. And as for not understanding our language, well, you can hardly fault us for that. If we are in the ivy tower, then you are surely in the basement. If we should descend so as to speak your language, why don't you ascend and meet us at least halfway up?'"

Thereupon the discussion would end in an impasse. The listeners would go away feeling that they had been led to the trough but kept from drinking, because the theoreticians had failed to say anything that made operational sense to them. "If applications are to be made," they would ask, "who is better able to make them than the minds that developed those ideas in the first place? It is only because they are uncooperative that we can't use what they have discovered."

We, the speakers, would go away equally disillusioned, feeling that we had been pouring the water into their open mouths but that they had refused to drink. "For who," we would ask, "should be better able to appreciate and apply what we have to say than the men who are daily involved with the very problems we have been analyzing? It is only because they are lazy and ignorant that they won't use what we have discovered."

And so, to point the moral, the uncooperative researcher-theoreticians and the lazy, ignorant practitioners would each go their own self-satisfied way, each convinced that the fault for any lack of communication lay with the other.

Now I recall also that when I made these speeches on the nomothetic-idiographic theme, I would usually start my remarks with the observation that I had never been an administrator myself and never hoped to be one. But about a half-decade after this time, I suddenly did find myself an administrator, not of a school system to be sure, but of a bureau of educational research and service, with a staff about the same size as might be found in a middle-sized school. I had the usual "honeymoon" and then my problems began. One day, perhaps six months after I had taken office, I suddenly sat up in my overstuffed administrator's chair and said, "Why most of my problems are being generated by people. People are no damn good!"

It was just at that point that the full significance of a farewell card that had been given to me by the staff associates in the Midwest Administration Center when I left Chicago hit me. "If you're so smart," it read, "why aren't you rich?" Or to quote from another idiom, the phrase "Physician, Heal Thyself" came home to me with a new forcefulness. How was it that a man of such great theoretical expertise in the staff relations area should suddenly conclude that people are no damn good? No use to claim that I couldn't understand the theoretical language -- I had helped to invent it![1] No use to claim that practice was hardly my concern -- I was up to my neck in it! What then was my problem? Why was I having so much trouble applying the ideas that I had myself helped to formulate?

The answer to that question was some years coming, and has two parts, I now believe:

(1) There is a tremendous gap between knowledge production and knowledge utilization that cannot be spanned either by the producer or by the utilizer himself, or even by these two acting in concert, at least in the typical situation. New mechanisms and agencies using special techniques are required to perform this bridging or linking function.

(2) Knowledge is at best only one of a number of input factors in any practical situation. No practical problem can be solved using knowledge alone -- a whole host of economic, social political, motivational, cultural, and other factors must be considered.

Let me illustrate these two points with some examples. First, in relation to the gap between knowledge production and knowledge utilization, education seems to be literally centuries behind other areas of endeavor in recognizing the gap and in making provisions for its reduction. In the physical sciences, for instance, engineering activities were instituted for precisely this purpose. Consider the Bell Telephone Laboratories as an agency for knowledge production and the Bell Telephone system as an agency for knowledge consumption and application. Now a great deal of knowledge production, commonly called basic research, goes on in the Bell Laboratories;

[1]The terms "idiographic" and "nomothetic" were picked from Roget's Thesaurus by me one wintry afternoon when I had nothing better to do than to try to find some new and interesting terms to use in our theory. We justified this at the time by claiming that we had to find terms "untainted" by value connotations.

to cite one instance, much of the research in solid state physics leading to the discovery and development of transistors was conducted there. But it is a long step from developing transistors as a laboratory curiosity, however exciting their potential might be, to utilizing the transistor principle in building better dialing and switching equipment. No one expects the scientists in the Bell Laboratories to make such applications; indeed, if anyone were to suggest it, the idea would be thrown out on the grounds that scientists would be diverted from what they did best and turned to a task that they could do but poorly.

Instead, AT&T in its wisdom has interposed a vast organization between the knowledge producers and the ultimate consumers. This system, known as Western Electric, has the unique mission of making the applications and producing the ultimate devices which the various Bell systems will install and use. Western Electric has its own coterie of engineers, who are themselves divided into specialties. Some of their personnel are concerned with developing prototype applications; others to testing these out and debugging them. Still others are concerned with designing these applications in ways that will make their production feasible and economical. And finally, of course, there are production specialists who actually turn out the devices that will be installed and used by the Bell Telephone companies.

This whole system seems to us only right and natural when we think of the physical sciences. But in education, even if there were good and plentiful basic research findings, there is no mechanism similar to the Western Electric Company, unless the R&D Centers and/or the Regional Educational Laboratories eventually assume this function to carry on the intermediate functions of development, testing, and production. And as my original example indicates, as recently as a decade ago this lack had never even crossed our minds; instead, we were content to write off the research-practice gap as stemming from the uncooperativeness of the researchers or the laziness and ignorance of the practitioners, or both.

Let me dwell now for a moment on the second part of my answer to the question of why there is so much difficulty in applying new knowledge, viz., that knowledge is, at best, one of a number of input factors in any practical problem situation. Let me use a real even if somewhat absurd example. In one school district I know about in the hills of Appalachia all of the power is held by the president of the Board who happens also to be the town physician. He has always controlled enough Board votes to hire and fire superintendents as he pleases. But this physician has one great vice: he is a morphine user. Now as a physician he had easy access to morphine and was able to provide himself with all that he needed to support his habit. But recently the state drug authority discovered his vice and relieved him of his license to prescribe narcotics. Hence he has had to turn to other sources for his supply, in this case, the local county health officer who is also his close personal friend.

Now it happens that the incumbent superintendent has somehow displeased the Board president, a failing that has cost the jobs of all of his predecessors. But the incumbent has one trump card: he happens to be the nephew of the county health officer. Hence the physician is faced with the difficult choice of firing the superintendent and losing his supply of narcotics or retaining the superintendent ahd having to put up with his nonconformist tendencies. A Hobson's choice indeed!

If we could find a candidate for the superintendent's job in this district who had ready access to a supply of drugs, great things might be accomplished. The incumbent is not in this happy situation, and every action he takes will have to be examined in terms of its potential for upsetting the delicate balance of power that presently exists. Get the physician too angry and he may decide that he can find some other source of drugs after all. If new knowledge is to be inserted into this school system's workings, it will have to be able to survive this scrutiny.

My colleague, Henry M. Brickell, has put the case more elegantly than this homely example illustrates. He says:

> When research-based information does exist, it must take its place beside all the other information available. The research finding may coincide with and confirm tha other information. In such a case, the chances of its being used are good. Or it may be the only source of information on a specific topic, in which case its chances of use are possibly only fair because it is not substantiated by experience. Or it may conflict with other information, in which case the situation is one of competition.
>
> In the United States even today, research findings do not compete well against such established, persuasive information sources as one's personal experience or knowledge of what other schools are doing. For example, when a local school asks, "What might we adopt to solve our particular problem?" a very limited number of solutions (at best) generated through a research and development process compete for its approval with a larger number of solutions which have been generated without benefit of research. The prospective adopter is not likely to select the research-based solution solely because it stands on a base of scientific knowledge, especially if something else is *less expensive, easier to install, preferred by the faculty, or otherwise attractive*. (Italics added.)[2]

---

[2]Henry M. Brickell, "Role of Research in the Innovation Process," Section V in Egon G. Guba (Ed.), *The Role of Educational Research in Educational Change: The United States*, a paper prepared for the Conference on the Role of Educational Research in Educational Change, UNESCO Institute for Education, Hamburg, Germany, July, 1967.

Let me call your special attention to the very last part of that quotation, which asserts that research-based solutions to educational problems are not likely to be selected if they are in competition with other solutions that are less expensive, easier to install, preferred by the faculty, or otherwise attractive. Mere knowledge, Brickell seems to be saying, is not enough; there are other economic, feasibility, and motivational factors that must be taken into account. And he might well have added social, political, cultural, and psychological factors as well. Whoever and whatever it is that will bridge the gap between knowledge production and knowledge utilization will have to be sophisticated enough and shrewd enough to assess these factors and be able to cope with them. In general I would assert that the typical researcher surely, and probably the typical administrator, do not have the special training and equipment for this purpose.

The Theory-Practice Continuum

If my analysis is correct, so that special mechanisms and agencies will be needed to fill the enormous gap between knowledge production and knowledge utilization, where are these to come from and what will their nature be? To deal with these questions I will need to digress for a moment to describe to you the categories of a theory-practice continuum which my colleague, David L. Clark and I have developed and published in other contexts.[3] I would like to begin by defining the various phases of this continuum and then proceed by discussing certain of their relationships.

[3]See for example our papers, "An Examination of Potential Change Roles in Education," NEA-CSI Seminar on Innovation in Planning School Curricula, Aerlie House, Virginia, October, 1965; and "Effecting Change in Institutions of Higher Education," UCEA International Inter-Visitation Program, Ann Arbor, Michigan, October, 1966.

Clark and I have talked about four phases or stages in this continuum, viz., research, development, diffusion, and adoption. Our concern today is with the middle two of these four, but I believe it is important to distinguish them from the other two, with which they are sometimes confused.

Research has as its basic objective the advancement of knowledge. The researcher is not concerned, nor should he be, with whether or not his research has an evident practical application. He needs freedom to pursue his ideas wherever they lead; he needs to be free to fail on occasion; he needs to be free from pressures for an immediate payoff. Research provides one input for the next phase, development.

Development has as its basic objective the identification of operating problems and the formulation of solutions to those problems. The developer, unlike the researcher, is most acutely concerned with practice. It is his job to make practice conform to the highest ideals that can be set for it, to be constantly probing the system to determine what, if anything, is keeping it from functioning at its best, and then to devise new approaches and techniques to ameliorate or eliminate whatever problems he may identify. In devising such problem solutions the developer borrows heavily wherever he can -- from research, from experts, from his own experience.

But development implies more than just coming up with an answer. The answer must be one that will work in the real world. It must be one that can be adapted into the system. It must be one that is usable by the personnel available. It must get results. Thus development involves production, engineering, packaging, and testing a proposed problem solution or invention.

Diffusion has as its basic objective the creation of awareness about new developments and the provision of opportunities for their assessment along whatever dimensions practitioners may deem necessary. The most potent solutions that men can devise to overcome their problems have little utility if practitioners are not informed about them, or if they have little opportunity to discover how the solutions work. Diffusion, in short, makes the solution available and understandable to the practitioner.

Adoption has as its basic objective the adaption of a development to the local situation and its installation therein. This is by no means an easy task. Every situation has its own peculiarities, so that it is unlikely that a newly developed problem solution, an invention, as it were, can simply be slipped into place without considerable modification to itself, to the system, or to both. Further, no prudent local administrator would agree to such an installation without some kind of previous trial. When the development passes this test there is still the matter of assimilating the invention as a component part of the system. This assimilation may involve the training of local personnel, modifying available space, arranging appropriate scheduling, and the like.

I have found it instructive, in thinking about these four stages, to develop a taxonomy of activities at each step that indicates what the researcher, the developer, etc., actually do. Again, we may consider each of the four phases in turn.

Research. It will be sufficient for present purposes to classify all possible research activities into four categories which I shall term depicting, relating, conceptualizing, and testing. This taxonomy, (Figure 1) is not generated in any systematic way but emerges from the

RESEARCH ACTIVITY
DEPICT
RELATE
CONCEPTUALIZE
TEST

Figure 1

following chain of reasoning:

When a researcher approaches a new topical area about which little is known, there is little that he can do other than describe the phenomena of interest. This description may take either qualitative or quantitative form. So for example a researcher might describe a group as being composed of both boys and girls, or as consisting of 67 per cent males. I shall use the term depict to refer to such a general description.

After a sufficient amount of depiction takes place it becomes possible for the researcher to relate depicted entities. So he may note that lipstick is worn exclusively by females, or that seven out of ten females wear lipstick while zero out of ten males wear it. He may also note that cancer of the lung seems to be related to cigarette smoking or that the correlation of height and weight is 0.71.

A sufficiently developed network of relationships makes it possible to suggest reasons for them. Why do certain phenomena tend to occur together? Why is lightening always followed by thunder? These questions lead directly to conceptualization, which we may regard as attempts to account for the observed depictions and relations.

These efforts at explanation may be tested to further determine the validity of the conceptualization. To the extent to which hypotheses are borne out, the formulation may be regarded as valid. In this testing process many of the same techniques used in the depicting and relating stages may be used again; typically, however, experimental methodology is employed which tests the hypothesis in a context-free (i.e., controlled) environment while holding the possible effects of other factors in abeyance.

The reconstructed logic of the research process is thus as follows: The aim of research is understanding. Understanding may be said to be achieved when a theory or taxonomy permits an explanation of the phenomena of interest, and of the relationships they bear to each other. Theories are built initially from systems of depictions and relations. The presence of the imperfect theory so devised makes possible more refined conceptualization. Further tests will confirm or deny the validity of the refinements. The four steps of depicting, relating, conceptualizing, and testing, successively repeated, will thus produce a very sophisticated science over time.

Development. Development activity may also be conveniently broken down into four categories which bear a curious similarity to the four categories of research. I shall term these development categories (Figure 2) depict, invent, fabricate, and test. They are derived by the same sort of intuitive logic as are the research categories.

DEVELOPMENT ACTIVITY

DEPICT

INVENT — Transmit, Translate, Transform, Synthesize, Create

FABRICATE

TEST

Figure 2

Development begins with the identification of problems. The developer is concerned with causing practice to conform to the highest ideals which he can imagine, but of course it never does. Certain desirable objectives are not reached, while other goals, perhaps even undesirable ones, are in fact attained. Those desirable goals which are attained may be achieved only imperfectly; there is always room for improvement. Whole new goals may become apparent for which the system makes no allowance, or older goals once considered important become less so. All of these factors require some alteration in the system. The developer's first job is thus to depict the state of affairs so that needs and problems can be identified.

Problems call out for solutions, and the developer's next task is to invent them. Now invention may take a variety of forms. First, it is conceivable that a solution already exists and simply needs to be applied. So, for example, a reading problem at the first grade level might be solvable through the adoption of the initial teaching alphabet (i/t/a). Perhaps a direct analog is known and simply needs to be adopted, e.g., teaching reading to blind children might be accomplished by adapting i/t/a to braille. Possible an indirect analog exists which can be converted into usable form, e.g., a reading program for teaching adult illiterates in the military might be transformable into a new introductory reading program for culturally disadvantaged youngsters. Or, the elements from which a solution may be devised may exist but may need to be appropriately combined to yield a solution; thus, several extant reading approaches may be combined to yield a relatively new approach. Finally, it may be necessary to invent a solution *de novo*, as was apparently done in the case of the initial teaching alphabet in the first place. We may speak of *transmitting*, *translating*, *transforming*, *systhesizing* and *creating* to describe these five different possible ways of arriving at a proposed problem solution.[4]

The fact that a solution is identified by whatever means does not signify that it is ready for application. Merely hitting upon an idea

---

[4]The three terms *transmit*, *translate*, and *transform* were coined by the Committee on Research Utilization of the American Educational Research Association to describe three ways in which research findings can be moved into practice. The terms have a somewhat different connotation here.

like i/t/a does not make it possible to begin using it at once. Materials have to be developed. These materials must be combined into appropriate sequences. The technique must fit into other ongoing school activities. I will call all of these operations taken as a whole fabrication; the term is intended to cover the entire gamut of engineering and packaging phases that may be required to make the invention "market ready," as it were.

Finally, the proposed solution must be field tested. It was devised to overcome some problem; does it in fact succeed? Does it work according to specifications? Should some refinements be made? Questions of this kind can be answered only through a comprehensive trial. And this trial must take place in authentic school situations; otherwise the applicability of the findings to the real world of education are dubious indeed.

The reconstructed logic of development is thus as follows: the developer, through a continuous monitoring of operational data (akin to process control), identifies particular operational problems which require solution. He invents a solution by transmitting, translating, or transforming already existing solutions, by synthesizing solutions from known but previously uncombined components, or by creating solutions de novo. In all of these processes he may look to research for guidance but research will be but one of several competing inputs. The invented solution is engineered into usable form, and finally is tested in a real school situation. It is then ready for warrant to the schools for use.

Let me digress here to make clear a fundamental distinction between research and development, two processes which are often confused. There are several reasons for this confusion. First we are often tempted to describe what I have here called "research" as "basic research," and what I have called "development" as applied research. This formulation gives the impression that research and development are simply different ends of the same continuum; indeed, someone has suggested that basic research is simply applied research with a time lag. But to commit this error is to ignore the fact that research and development have entirely different objectives; they are complementary processes to be sure, but they serve different goals.

A second reason for the confusion is that persons engaged in research and development often are seen to be using similar techniques. Thus similar instruments, design, field procedures, and data processing methods may be observed. But surely we will not fall into this trap; to do so would be akin to saying that because plumbers, carpenters, and masons all use hammers that they are all doing the same thing.

A more pervasive and compelling reason for confusing research and development stems, I believe, from our intuitive understanding that the gamut of activities embraced by each tend to begin and end in analogous operational modes, just as our taxonomies of research and development both begin and end with the same terms: depict and test. I have juxtaposed the two taxonomies in Figure 3 to make this clear.

RESEARCH ACTIVITY	DEVELOPMENT ACTIVITY
DEPICT	DEPICT
RELATE	INVENT
CONCEPTUALIZE	FABRICATE
TEST	TEST

Figure 3

The crucial differences between these two phases may be delineated by going back to the basic purpose or objective of each activity. The researcher depicts much as a painter depicts: he attempts to discover the salient elements in the situation and then to portray them in their appropriate relationships and contexts. The developer depicts not to portray the process but to monitor it; to discover problems in their still incipient stages and thus to be able quickly to counteract them.

The researcher tests in order to verify or refute his hypotheses. It is imperative in his testing that he maintain rigorous control over all elements so that only those that enter specifically into the hypotheses can interact. It is in this way that we investigate the law of gravity for example, and can show, under conditions of a vacuum, that a feather and a stone do indeed fall at the same rate. Thus we establish universal laws. The developer is not concerned with controls, however. He does not need to know what happens to a stone and a feather under idealized conditions but in the real world. When he develops a solution to a problem it must be clear that it will work not only in the best of all possible worlds in which everything irrelevant can be constrained but also in the worst of all possible worlds in which everything irrelevant is free to contaminate. We shall return to this problem in our later discussion of evaluation; for the time being let it suffice to demonstrate that the testing of the researcher is not different just in degree or time from the testing of the developer but in fundamental intent.

Diffusion. The activities in which a diffusion agent, or diffuser, engages are those that are involved in bringing a proposed problem solution or invention to the attention of some one who may actually use it in practice, and those involved in giving that practitioner the opportunity to assess the operating qualities of the invention. A taxonomy appropriate to this range of activity is shown in Figure 4. There seem to be essentially six ways in which the diffuser may operate:

DIFFUSION ACTIVITY
TELL
SHOW
HELP
INVOLVE
TRAIN
INTERVENE

Figure 4

1. He can tell. Telling involves the word. The word may be written, as in newsletters, papers, monographs, books, articles, and the like; or it may be spoken, as in conferences, speeches, conversations, etc. My essential diffusion mode today is, obviously, telling.

2. He can show. Showing is a form of communication which involves a direct confrontation with the phenomena of interest, as in a planned or casual observation, or in actual participation. It may involve structured experiences such as demonstrations or simulations; or it may involve looking at materials or displays such as pictures, slides, films, dioramas, realia, and the like.

3. He can help. Helping consists in the direct involvement of the diffuser in the affairs of the practitioner but on the practitioner's terms. It may take the form of consultation, service, trouble-shooting, and the like.

4. He can involve. Involving takes the form of an inclusion or cooptation of the practitioner. Thus the diffuser may enlist the practitioner in assisting with the development, testing, or packaging of an innovation; in acting as a "satellite" or agent to diffuse the invention to others; in contributing the problems to which innovative solutions are to be sought; and the like.

5. He can train. Training takes the form of familiarizing practitioners with the features of the proposed problem solution or invention, or of assisting them to increase their skills and competencies or to alter their attitudes. It may be accomplished through formal university credit courses, institutes, workshops, internships, apprenticeships, extension courses, local in-service training, "T-sessions," and similar experiences. Training may involve telling, showing, helping, and involving but differs from these other techniques in that the practitioner makes a formal commitment to learn by allowing himself to be trained.

6. He can intervene. Intervening consists in the direct involvement of the diffuser on his own terms, not those of the practitioner. It may take the form of mandating certain actions (e.g., adopting a statewide textbook), inserting certain control mechanisms (e.g., instituting a statewide testing program), or of intruding certain economic or political factors (e.g., arranging the purchase of language laboratory equipment or causing board dismissal of an uncooperative teacher).

The reconstructed logic of the diffusion process is thus as follows: The diffuser has the task of building awareness and understanding of an invention and causing practitioners to consider its features with a view to possible application. To discharge this function he has essentially six techniques at his disposal: telling, showing, helping, involving, training, and intervening. He will use any combination of these techniques to cause favorable consideration without resorting to hucksterism or unethical manipulation. He sees himself as a person opening viable professional alternatives to the potential adopter with a problem to solve.

Adoption. The purpose of adoption activity is to shape and install a problem solution or invention within a particular local setting. This phase seems to have received little conceptual attention from anyone; it is perhaps the most muddy of the four. It seems to me that at least three major steps are involved, with the second of these being divided into several sub-steps (Figure 5) as follows:

ADOPTION ACTIVITY

TRY-TEST

INSTALL---------- Modify / Train / Equip and House / Organize

INSTITUTIONALIZE

Figure 5

1. Trial. No prudent administrator will permit the installation of a proposed problem solution on a permanent basis without having convinced himself that it will perform as claimed. Indeed, a local trial is mandatory even when national assessments have indicated that the solution performs well on the average, for the obvious reason that the situation in which installation is proposed may not be average. Local variations must be taken into account.

2. Installation. When a proposed solution has proved itself through a local trial, it then becomes necessary to arrange for its installation on a building wide or system wide basis. At least four areas of concern must be attended to:

a. Modification. No invention will fit exactly into a local school situation for which it was not explicitly designed. Decisions will have to be made whether the fit can best be accomplished by modifications in the invention itself or in the school situation. If for example the invention requires teachers with particular skills but teachers with these skills are simply not available, some modification in the invention will be required.

b. Training. Personnel expected to use the invention must be trained. No teacher will willingly risk his reputation before a class with a technique about which he is unsure. More importantly, no administrator should be willing to permit a teacher to adopt a new technique without proper training for use, lest through lack of knowledge he should fail to take full advantage of whatever additional benefits are expected to accrue.

c. Facilities. Many inventions require particular kinds of physical arrangements. Typically a school adopting such an invention will not be suitably housed for the purpose or may not possess appropriate equipment. Flexible scheduling or multiple-size grouping cannot occur in a building arranged for conventional size classes of 25 or 30.

d. Administration and organization. The proposed invention may have important administrative or organizational consequences. Problems in scheduling, in budgeting, in staffing, in organizing may all produce headaches for the administrator. Unless these possibly disruptive consequences can be foreseen and obviated, the result may be a failure of an otherwise useful invention.

3. Institutionalization. Ultimately the invention must be assimilated into the ongoing program. At some time it must cease to be viewed as new and must become an integral and accepted component. It is not clear to me what steps might be taken to insure institutionalization. Sometimes I feel that the most important factor may simply be the passage of time.

obviously, the lack of awkward incidents in relation to the invention is helpful and the more quickly the spotlight can be taken off of it the more quickly it is likely to become accepted.

How Are We Doing on Bridging the Gap Between Research and Practice?

In my preceding remarks I have attempted to illustrate the fact that there is a large gap between knowledge production and utilization, and I have attempted to depict the flow of knowledge from initial research into final use in terms of a four-category continuum. Our concern at this conference is primarily with the middle two categories, development and diffusion, for they represent the projected means for bridging this gap. I would like to turn now briefly to a consideration of what we are doing, and how well, in operationalizing these two categories.

I will, therefore, not make any further remarks about either research or adoption. I do feel compelled to observe, however, as we leave these categories behind, that my lack of attention to them does not indicate any high degree of satisfaction on my part with the way research and adoption activities are operating. Indeed, it is well known that research results are not being utilized to any great degree in educational practice, and that almost no attention has been paid either conceptually or practically to the problems of adoption which I briefly outlined above. But my concern today is with the bridge and not with the abutments, although I hope that due attention will be paid lest we mount our bridge of steel on banks of sand, when the time comes.

Let me turn then to a more detailed consideration of development and adoption.

Development. Development is a very complicated process which neither practitioners nor researchers are particularly competent to carry out. If there is any area in education that calls for reorganization and for the evolution of new professional roles, this is certainly it. Experience from industry indicates that from five to eleven times as much investment is required to develop an application from a research finding than was necessary to produce the research finding in the first place. High level specialists are required to do the job. Moreover, development depends not only upon the availability of relevant basic research but upon a host of other factors as well: the availability of resources, institutional support, experience, practical judgment, political factors, and the like. Research data provide only one of several critical inputs, and the blending of these inputs requires more specialized skill than either researchers or practitioners commonly possess.

Initial attempts at development in education occurred gradually and without a clear realization of what was happening. I am sure that the persons following the lead of Jerrold Zacharias in the development of the PSSC physics materials were scarcely aware of what a vanguard group they were. The several other curriculum development groups, mainly funded by the National Science Foundation in those early days, were certainly more interested in updating content than they were in establishing development patterns which others might emulate. But their pattern did seem to prove successful, and it was soon emulated, particularly in the new course content improvement projects of the U. S. Office of Education.

In more recent months we have seen further systematic attempts to establish development agencies. Clearly the research and development centers have a mandate to turn their research into practice. But as we have seen, successful development involves a great deal more than the mere availability of relevant research. We may well wonder therefore whether the primarily research oriented R & D centers will be up to the task. Another similar effort has occurred in the establishment of the regional educational laboratories, which are mandated to identify and solve educational problems, hopefully through recourse to research but by other means if necessary. Thus far the laboratories are too new to make it profitable to venture a judgment about their probable level of success.

It seems that no existing agencies have responsibility for the full range of development activities indicated by the taxonomy presented earlier. The depicting function seems to be especially neglected. While both regional laboratories and Title III projects were mandated to make needs surveys of their regions, it is clear that these surveys were carried out in a most perfunctory way, and without the benefit of hard data in many cases. (I should note at once that this is not the fault of the agencies involved so much as it is of the Office of Education, which mandated these surveys under incredible constraints of time and resources.) More importantly, even when well done, these surveys provide but a static "snapshot" of the situation at any moment rather than a dynamic "motion picture film" over an appreciable time span.

The invention function is perhaps better managed than the others, although certainly not nearly as well as it should be. Funds are available for improvement projects and several agencies, including the new industry-education combines as well as the regional laboratories and research and development centers, are beginning to undertake massive improvement projects. Yet a conceptual underpinning for such activity is still missing. We still know far too little about effective ways of creating new solutions or even of transmitting, translating, or transforming known solutions.

Fabrication will probably be handled best by the industry-education combines, since these typically involve publishers and manufacturers of hardware that can be used to good effect. The publishing industry has shown a great deal of ingenuity in the past in placing its materials into interesting and novel formats and will probably continue to do so.

In the area of testing we come again upon a quite underdeveloped area. We shall see later that existing evaluation designs do not seem to be too appropriate for the real problems of education. We may also be concerned that if much of the fabrication is carried on by commercial agencies, they may be over eager to rush their fabrications into production without the kinds of testing that would assure a professionally warrantable product. Thus both conceptual and consumer protection innovations are needed in the area of testing.

From one point of view, then, the development picture is not too rosy. When one considers, however, how late in the day we determined to undertake development at all, and with what meagre resources we have supported it, we may perhaps be forgiven if we take a more charitable view. Now that education is fully aware of the need for development activities, is apprised of their complexity, and is being aided with resources to get development activity started, we may hope that within a decade most of the problems I have enumerated will have disappeared.

Diffusion. Diffusion is an activity regarded with some distaste by many members of the educational establishment, particularly the research community. It is often equated with hucksterism, and I suppose, in fairness, that one must concede that a great deal of hucksterism does take place. This fact may be the best argument one can muster in favor of well organized diffusion efforts, however, so that one can be sure that what is being diffused is a viable alternative rather than just another fad.

Traditionally educational diffusion has fallen within the domain of commercial interests, mainly the book publisher. Recently both research and development centers and regional educational laboratories were given some diffusion responsibilities, and these agencies have begun to develop new approaches, although haltingly.

The major diffusion responsibility seems to be falling squarely on the shoulders of Title III projects. There is a school of thought that suggests that research and development centers should be concerned with research, regional educational laboratories with development, and Title III projects with diffusion. This is a formulation with which I am in essential agreement, perhaps because this division of labor would fit my earlier model so well. There would be at least three of the change stages, then, for which institutional responsibility would be firmly fixed. This formulation also seems to be supported in the Office of Education.

But whatever our view may be about the appropriate institutional arrangements for carrying out the diffusion function, it is clear that so far that function has not been carried out very well. In my own opinion the major reason for this failure may be traced to our earlier failure to delineate acceptable strategies for diffusion. I use the word strategy to indicate an action plan which indicates which of the adoption techniques outlined in the earlier adoption activity taxonomy should be used when and where and in what combination. To evolve such a strategy seems to me to imply some consideration of at least the following elements:

1. Assumptions concerning the nature of the practitioner who will be exposed to the strategy. The practitioner may be viewed as a rational entity, who can be convinced, on the basis of hard data and logical argument, of the utility of a proposed invention; as an untrained entity who does not know how to perform but who can be taught; as a psychological entity who can be persuaded; as an economic entity who can be compensated or deprived; as a political entity who can be influenced; as an entity in a bureaucratic system who can be compelled; or as a professionally oriented entity who can be obligated. We might term these respectively as rational, didactic, psychological, economic, political, authority, and value assumptions. Obviously the ways in which the earlier outlined techniques are used will depend heavily on which assumptions one makes. So for example, telling, showing, training, etc. will certainly be different if one assumes a rationally oriented subject (i.e., one who will be convinced by facts) than if one assumes a politically oriented subject (i.e., one who can be manipulated).

2. <u>Assumptions concerning the end state in which one wishes to leave the practitioner</u>. Very little attention is typically paid to the question of the end state in which the diffuser wishes to leave his subject. This situation may arise, of course, because the diffuser may act as a mere huckster; hucksterism may "sell" a particular invention being promoted but it may leave the practitioner with very little residual propensity ever to consider any other proposed invention. But even with "well-intentioned" diffusers this difficulty may arise because of a basic failure to consider desirable end states. What is it that the practitioner should be able to do, think, or to feel as a result of having been exposed to a diffusion strategy? Is he to be better trained? More skillful? More knowledgeable? More open? Wiser? Obviously the choice of a diffusion strategy would be considerably aided by careful attention to this factor. It seems particularly ironic that this situation of carelessness about end states should hold true in the field of education, which is so generally characterized by concern about behavioral outcomes and objectives. If we applied a little of our usual logic about specifying expected goals this difficulty would be largely overcome.

3. <u>Assumptions about the nature of the agency or mechanism carrying out the diffusion activity</u>. No sensible diffusion strategy can be evolved without careful attention to the matter of who is to carry it out. For not all strategies are within the capabilities of all agents or mechanisms. Constraints exist which mandate certain actions for certain agents and which prohibit other actions to them. So for example, a regional educational laboratory, acting as a diffusion agent, is hardly in a position to intervene,

since it lacks the necessary power or authority to do so, but telling, showing, or involving come "naturally" to it. A state department of education may well intervene (and indeed may be legally mandated to do so) but would probably be very suspect if it tried to involve. An individual teacher can tell and show but probably would be thought ridiculous if she set up a training experience for her fellows. A university, however, could carry out this latter function with impunity. Since the final implementation of the strategy depends upon the agent, the strategy must be one appropriate to the agent's circumstances.

4. Assumptions concerning the substance of the invention. Obviously not all inventions are alike; they pose different problems of adoption, and this fact must be taken into account in developing an appropriate diffusion strategy. One way to view this situation is in terms of the amount of change mandated by the invention. Thus Chin characterizes innovations as involving mere substitution, alteration (a minor change,) perturbations and variations (mere changes in organizational equilibrium), restructuring (requiring reorganization), and value orientation change (deep-seated value changes).[5] Rogers talks about characteristics of inventions that make them more or less acceptable, including relative advantage (intrinsic superiority), compatability (consistency with existing values and experience), complexity (difficulty in use), divisibility

---

[5]Robert Chin, "Models and Ideas About Changing," paper presented at the "Symposium on Identifying Techniques and Principles for Gaining Acceptance of Research Results of Use of Newer Media in Education," W. C. Meierhenry, Sumposium Director, Lincoln, Nebraska, University of Nebraska, 1963.

(degree to which the invention can be partitioned and/or tried on a limited basis), and communicability (or diffusability).[6] Whether these or other ways of classifying the substance of innovations are most useful is less important for us at the moment than that there be some explicit way for taking account of substance at the time that a diffusion strategy is devised.

We are thus confronted, in considering diffusion, with a picture that is, if anything, even less satisfying than that presented by development, which we reviewed earlier. There seems to be a considerable confusion about the organizational responsibilities that may exist in this important arena, with attempts to develop viable organizations being so recent as to invalidate any attempts at judgment at this time. Further, theory and practice are both relatively silent on the important issue of how diffusion strategies are best devised. All we seem to be able to do at this time is to point to the important factors that probably ought to be considered. However, as in the case for development, when one considers how recently this concern has emerged and how new are our efforts to deal with it, we may perhaps be willing to take a more long range view.

Evaluation

Thus far I have said very little about evaluation, which you may have considered rather remarkable in view of the fact that the term appears in the title of this paper. I wish to remedy that defect now. Evaluation is to important and so pervasive a concept when we think about closing the

---

[6] Everett M. Rogers, Diffusion of Innovations, Glencoe, Illinois: The Free Press, 1962, p. 124.

gap between knowledge production and utilization that it deserves quite detailed and separate attention.

I shall have two major points to make about it: (1) the concept of evaluation is changing rapidly, becoming in particular much more pervasive than has traditionally been the case, and (2) the methodologies currently in use for evaluation are hopelessly bad and urgently need replacement.

Let me begin with some observations about what has in the past been meant by evaluation. Typically two complementary operations are denoted by the term: (1) the comparison of some results, output, or product with a set of standards, in an absolute sense, and (2) the comparison of some two or more methods of producing the same results, output, or product, in a relative sense. In the first case the standards were usually derived in relation to some objective. Thus, the objective might be to develop reading skill, and the standard might be the 4.0 grade equivalent on the Stanford Reading Achievement Test. Pupils could then be judged, in an absolute sense, on their achievement of that objective. Or two methods of teaching reading skills might be judged to determine which produced a higher average reading skill level in two groups of pupils, in a relative sense.

Measurements taken to carry out these classic forms of evaluation are usually of the pre- and post-test type, depending upon one's preoccupation with initial status, group equivalence, and similar matters relating to control or data analysis. The term <u>bench mark</u> is frequently used to describe collection of initial status data. Between collection of bench mark data and final performance data a long period, say a semester or school year in length, could and usually did intervene, during which data might or might not be collected but during which stringent controls are maintained so that the data will not be confounded. In particular great care is taken not to alter any essential element related to the method, technique, or content being evaluated, lest the change render the evaluation invalid (one could not tell what was being evaluated). Generally speaking the traditional rules of experimental design and field control are rigorously invoked. The essential task of traditional evaluation is to judge.

Emergent evaluation however is seen as a tool to aid in decision-making. The tasks of (1) identifying an educational problem or need, (2) devising or selecting a treatment to cope with it, (3) implementing the treatment procedures, and (4) determining the treatment's feasibility, quality, effectiveness, and efficiency require a series of decisions which evaluation can aid. The process of collecting and interpreting data relevant to this series is seen as the substance of evaluation.

Daniel Stufflebeam of the Evaluation Center at The Ohio State University seems to me to have come closest to defining the new evaluation when he talks about four kinds of evaluative activity. The first of these is context evaluation, which, in the setting of the school, means the continuous determination of the school's status on key variables with a view to identifying needs and problems. Such an evaluation gives the decision maker data he needs to have about important directions in which he should move. Second, there is input evaluation, which is concerned with assessing various possible responses to the needs or problems that may exist. There are probably a number of ways, for example, in which a school principal might revamp his reading program to take account of the special problems posed by culturally disadvantaged children; which of these ways has the highest payoff potential in his situation? Third, we need to be concerned with process evaluation, which is used to determine whether the selected input is working as it was expected to and which, even more importantly, provides for continuous feedback so that the selected input can be continuously refined and adjusted to better achieve its intended purpose. Finally, there is product evaluation, which is most like what we have traditionally meant by evaluation, i.e., the determination of the feasibility, quality, efficiency, and effectiveness of the input in responding to the need or problem involved.

It is interesting to check the terms of this analysis against the terms listed in the taxonomy of development presented earlier. What Stufflebeam calls context evaluation is of course very similar to what I meant by the term depict, i.e., a continuous assessment of the situation. We might note the similarity of this concept of continuous assessment to the older concepts of bench mark or base line, but while these latter are static concepts indicating status at some point in time like a snapshot, the continuous assessment idea is rather like a dynamic bench mark or base line, giving, as it were, a continuous motion picture film of what is going on. Needless to say attempts at continuous assessment pose some interesting methodological problems.

Next, it seems clear that Stufflebeam's idea of input evaluation has relevance at what I have called the "invent" stage of development.

In order to determine, for example, whether the invention problem is one of transmitting, translating, or transforming existing solutions, of synthesizing new solutions from available elements, or of creating a solution de novo, some assessment will be required of possible inputs and their probability of useful payoff.

Finally, when a solution has been fabricated, it must be tested, and it is clear now that testing should involve both process and product measures. It is likely that the solution will not be in near-perfect form when it is first applied in a real context; hence continuous improvement is mandated. Process evaluation allows for this contingency. Further, we need to be sure that the solution is being applied in a form reasonably similar to the one its fabricators had in mind; again, process evaluation to the rescue. And of course we want to be sure that the solution does in fact achieve its objectives; i.e., meeting the need or responding to the problem. And here we have product evaluation.

Needless to say, we are a long way indeed from having the techniques necessary for applying evaluation in the way indicated by this analysis. These concepts are only now emerging, and it will take a long time before we are able to apply them systematically in operational situations. But it is clear that traditional concepts are no longer good enough.

The shortcomings of traditional evaluation can be documented in other ways than through such a theoretical analysis, however, We need only to look at the large mass of "no significant difference" findings typically produced by evaluation studies to begin to wonder about the power of the techniques, particularly when all the evidence of the senses of participants argues that there is a difference. Or consider the conclusion of the widely-publicized Coleman report, which asserts, after a most careful and thorough examination of all available data, that there is only a ". . . relatively small amount of school-to-school variation that is not accounted for by differences in family background, indicating the small independent effect of variations in school facilities, curriculum, and staff upon achievement."[7]

This conclusion is simply incredible on its face. It means, if true, that it makes no difference whether a teacher is good or bad, whether good or poor materials are available, whether the school is a barn or a geodesic dome, students will learn about the same (and not much at that!). Now anyone who has spent any time at all in a school knows that is just not so; why then do our evaluative techniques not pick this up?

I believe it can be argued that traditional evaluation has four characteristics which account for its sharply limited utility. These include <u>terminal availability of data</u>, <u>retrospective view</u>, <u>imposition of constraints</u>, and <u>limited generalizability</u>.

1. Evaluative data are usually available only upon the termination of the evaluative period. Hence they can provide information relevant only to "go," "no-go," or "recycle" decisions about the treatment being evaluated. Other kinds of decisions cannot be served.

---

[7]James S. Coleman, <u>et al.</u>, <u>Equality of Educational Opportunity</u>, National Center for Educational Statistics, U. S. Government Printing Office, Washington, D. C., 1966, p. 235.

2. Evaluative data typically afford only a retrospective view. The evaluation does not provide information during the test of the treatment which might have been used to improve it.

3. The assumptions on which evaluative designs are based (those of traditional experimental design) impose a series of constraints on the evaluator. There can be, for example, no variation in treatment or context once the evaluation is under way, since this would result in the confounding of critical variances. Thus traditional evaluations militate against any concurrent effort at improvement of the treatment and against other contextual changes, e.g., the introduction of any other innovation, during the term of evaluation.

4. The constraints imposed because of the requirements of classical experimental design in effect create a laboratory-condition within which the treatment will be tested. The many sources of variation found in the real world are deliberately excluded from having any effect upon the outcome. The evaluation describes what happens under laboratory circumstances, and not under "typical" circumstances. The generalizability of the findings is thus necessarily limited.

The problem of constraints is an especially interesting one and probably deserves some special comment. Generally speaking, the constraints arise because of a variety of assumptions that must be made to support the logical and statistical structure of design theory. Three general classes of assumptions may be identified:

1. Statistical assumptions. Statistical assumptions support the development of the statistical techniques for analyzing and interpreting

data. There are certain assumptions necessary to know that a distribution is normal before one can assert that 68 per cent of the cases are included in the interval $\bar{X} \pm s$. Other assumptions are built into the derivation of the interpretive tables in which the "significance" of analytic statistics is read; thus the derivation of the F distribution depends upon assumptions of random sampling from a population in which the variable of concern is normally distributed. Finally, still other assumptions are necessary to support the logic of an analytic method. Thus, in the case of analysis of variance (and other tests of significance), the additivity assumption which asserts that treatments have equal effects on all persons to whom they are applied, is vital. For unless this assumption is met, group variances change and the basis for computing an error term disappears.

2. Design assumptions. A second class of assumptions has to do with the logical requirements of design procedures. Typically, in an experiment, the effect of some treatment is to be determined. A group exposed to the treatment cannot simply be measured to determine that effect, for the obvious reason that there exists no "bench mark" against which to assess that measurement. If the bench mark is provided by simply making a second (earlier) measure on the experimental group, the difference may still be called into question as having been caused by other extraneous(confounding) effects such as history, maturation, and the like. A second group, the control group, is usually added to obviate this difficulty. But the second group is useless unless it is comparable to the first; because design procedures have been worked out on the

assumption that such comparability exists. The function of the assumption of comparability is to protect the internal validity of the experiment.

But external validity (generalizability) may be threatened also. First, we need to be certain that experimental and control groups were in fact selected in some way that guarantees their representativeness of the population to which the results are to be generalized. Under ideal circumstances such representativeness can be guaranteed only by both random selection of subjects from the population and random assignment of the subjects to the experimental and control groups (random selection and assignment would of course also guarantee comparability). Then, we must protect the groups against reactive or interactive effects that would alter the groups in some way during the experiment so that their representativeness would be dubious.

Thus the requirements of design theory as it is now explicated require comparability to protect internal validity, and random selection, random assignment, and reaction-interaction control to protect external validity.

3. Treatment assumptions. Statistical and design assumptions are quite well understood because these assumptions had to be made explicitly in order that the statistic or the design could in fact evolve. Less well understood are the implicit assumptions made about the treatment whose effect is to be tested. It must be assumed that the treatment is fully explicated a priori. It must be assumed that the treatment can be "plugged into" the experimental setting with no interactive effect with

other elements. The treatment must further be invariant throughout (else the variances are confounded) and must be applied in identical ways by all persons responsible for its trial. Finally, it must be the case that there are no competing treatments, for if such competing treatments exist their individual effects can not be separated.

All of these assumptions are in some particulars unrealistic for education. Among statistical assumptions, for example, the additivity assumption is especially inappropriate; every experienced teacher knows that effective teaching will increase the variance of the group being taught, and usually markedly. Among the design assumptions the comparability problem is especially sticky. Usually comparability cannot be managed directly. So an indirect process, such as locating schools with similar buildings, similar socio-economic backgrounds, similar intelligence levels, etc., is used. Such procedures may or may not solve the comparability issue, but they certainly do destroy external validity, at least to the extent of limiting the generalizability of the findings to similar restricted groups. Finally, among the treatment assumptions, treatment invariance is not only quite difficult to achieve, but may be undesirable, since the treatment may be one that could profit from continuous improvement even while being tested.

This analysis has led me to the conclusion that some new evaluation strategy free of the defects that I have enumerated is necessary before evaluation as a science can make its next major strides. Of course no such strategy exists at the moment, but it is possible to indicate certain characteristics which it must have if it is to be successful:

1. Level of control. Typical experimental controls must be eliminated. The evaluator must be concerned with how things occur in the field rather than in the laboratory. Hence the kind of control that we have been accustomed to in laboratory experimentation will be sharply different, perhaps non-existent.

2. Intervention. The evaluator cannot arrange the inquiry situation but must accept is as it occurs. Data collection must be carried on in a non-intervention mode, i.e., without disturbing either the context or the subjects.

3. Continuity of data collection. Data are not collected simply at pre- and post-experimental periods (or at some particular check points) but continuously throughout the evaluation. The baseline of data must be dynamic rather than static.

4. Treatments. Treatments cannot be regarded as invariant but as susceptible to continuous change (improvement). Context conditions must also be alterable.

5. Scope. Attention must be given not only to particular variables which have been identified and operationalized beforehand, but to any emergent variables which appear to be of concern.

6. Assumptions. The evaluation system cannot be cause to conform to traditional assumptions, but rather the assumptions must be formed to meet the reality of the situation. It is only on such reality-oriented assumptions that a useful theory of evaluation can be based.

Conclusion

Well, as you can see, I haven't learned much over the last decade; I am still playing the old nomothetic-idiographic game in a new guise. However sparkling my ideas may be, they certainly do not provide very much operational guidance. But that is exactly the problem I have tried to deal with, and this presentation is as good an example as any of the fact that there is an enormous hiatus between theory and practice.

Whether you agree with my particular formulations or not interests me less than that you agree that there is a problem of fantastic proportions confronting us. It is a problem that will not be solved without a great deal of attention to emergent roles and organizations. We have made some strides on the problem of organizations, as witness the new agencies and programs that have emerged as a result of the passage of ESEA. On the personnel problem we are still far behind; so far as I know there is no program in existence anywhere making a concerted effort to train the range of middlemen that will be required. Indeed, we are still lacking even a primitive formulation of the roles that should be created. We certainly do not know where we shall recruit the persons who will ultimately fill those roles. We have no materials with which to train them.

Where shall we turn for responsible leadership? It seems to me that leadership must, at this moment, come from the two existing establishments that are necessarily most concerned about the gap--the educational researchers, who stand at one end of the knowledge production-utilization continuum, and the educational administrators, who stand at the other end and who bear the responsibility for effective practice. Neither group, I am convinced, can do this job alone, and neither group ought to attempt to do it alone. But both groups must cooperate to get things started.

I don't know how the initial step should be taken. Perhaps a national commission of researchers and administrators should be appointed, with or without the blessing of the U. S. Office of Education. Perhaps AASA and AERA should combine in this venture. Perhaps university faculties in these two areas should develop joint programs for the training of middlemen. Perhaps UCEA can serve as an appropriate forum. Whatever the route, the action must come soon, for the need is great.