DOCUMENT RESUME

ED 088 844 SP 007 815

AUTHOR Medley, Donald M.
TITLE Indicators and Measures of Teacher Effectiveness: A Review of the Research.
PUB DATE [71]
NOTE 29p.

EDRS PRICE MF-$0.75 HC-$1.85
DESCRIPTORS *Educational Research; *Effective Teaching; Evaluation Techniques; Literature Reviews; Measurement; *Measurement Techniques; Research; *Teacher Evaluation

ABSTRACT

In this speech the author considers the problem of measuring teacher effectiveness from the point of view of an educational researcher. He examines two prevalent methodological approaches: evaluating the process and evaluating the product. The first approach involves the description of a study which concluded that teachers rated high by supervisors and well-liked by pupils were not the teachers whose pupils showed the greatest gains and who judged themselves most effective. The second approach reviews the problems of the reliability and validity of tests and advocates giving up the idea of measuring teacher effectiveness in terms of pupil gains on tests. The recent work of Flanders, Bellack, Smith, Hughes and others is commented on, with the conclusion that the more dimensions of teacher behavior that can be measured and correlated with pupil gains, the more likely is the discovery of one or more correlating with teacher effectiveness. Reviewing common practices that measure teacher effectiveness, the author points out that educational research does not support Dewey's "we learn by doing," or Skinner's reinforcement theories. Instead, less obvious things, such as the cognitive level of questions asked, probing questions, clarity and structure, should be looked for in measuring teacher effectiveness. (JA)

INDICATORS AND MEASURES OF TEACHER EFFECTIVENESS: A REVIEW OF THE RESEARCH

Donald M. Medley
Bureau of Educational Research
University of Virginia

As a research worker I am a bit embarrassed at being asked to speak to this group on this topic. It is not that the topic is outside my area of interest -- on the contrary, I have devoted my professional life in large measure to this very problem of trying to findout what teacher effectiveness is and measure it. And it is not, as some of my critics might suggest, because I have had so little success that I had rather not discuss it. My discomfort relates rather to a difference in the concerns and constraints which govern the way you see the problem and those under which I see it.

You no doubt remember what old Mrs. Murphy said when young Miss Reilly remarked on the way home from Sunday mass:

"Ah, 'twas a lovely sermon Father O'Toole gave this morning on the joys of motherhood."

"Indeed it was," said Mrs. Murphy with a sigh. "I only wish I knew as little about it as he does."

To the research worker the problem of measuring teacher effectiveness is no different from any other problem, except, perhaps, that it is more complex than most, and cannot be studied in the antiseptic environment of the laboratory. The nice thing about research is you don't really have to succeed. If I as a research worker try to measure teacher effectiveness and fail,

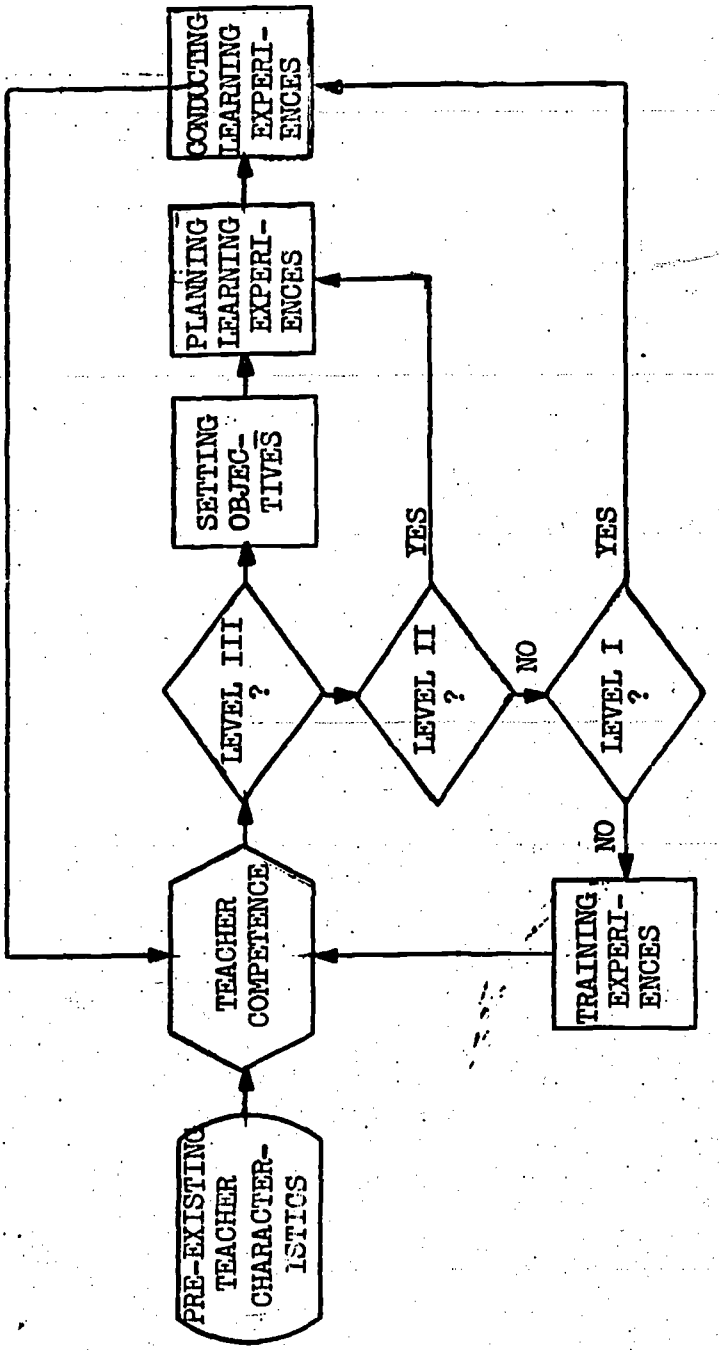

EXHIBIT 1

I publish the study anyhow and try again.

The practicing educator is in a very different situation. For one reason or another he must evaluate teachers whether he can measure their effectiveness or not. So I guess I must not--do what I feel best qualified to do today, and that is tell you we do not yet know enough about the nature of effective teacher behavior to be able to measure it--yet. If you can wait ten years, or maybe twenty, we may be in a better position. But as of now--well, my best advice would have to be--don't try.

But you must try, and succeed. So I am going to stick my neck out somewhat, and try to summarize for you the best guesses I can make--and the word guess is the appropriate one to use--about how to go about this important task of measuring teacher effectiveness.

Before I start, may I suggest that we accept, at least for today's discussion, a clarification in terminology between three words often used interchangeably: skill, effectiveness, and competence. (Exhibit 1)

In studying teacher behavior and research related to it I have found it useful to distinguish them basically in terms of measurement strategy--in terms of what task you would set a teacher to measure each of the three.

To tell whether a teacher is competent, I would give a teacher a class and say, "educate them." The teacher would need to define appropriate objectives for the pupils, plan ways of achieving those objectives, and execute the plan. In order to assess competence you would have to measure the quality of the

objectives chosen, the appropriateness of the plans, and the ability of the teacher to execute them.

To tell whether or not a teacher is effective, I would give a teacher a set of objectives and a class and say, "achieve them." The teacher would need to plan how to achieve the objectives and execute the plan. In order to assess effectiveness you would have to assess the appropriateness of the plan and the ability of the teacher to achieve it. You would not need to judge the quality of the objectives.

To tell whether or not a teacher is skillful, I would give a teacher a plan and a class and say "carry it out." The teacher would need only to know how to execute a plan--conduct the discussion, operate the hardware, or whatever. In order to assess skill you would need only assess the teachers ability to do these things successfully. You would not need to judge the quality of the objectives or the appropriateness of the plan, since these were given.

I was invited to come here today and talk about research in measuring teacher effectiveness. I am not sure about it, but I am going to assume that the term effectiveness was used in the sense in which I have just defined it--as referring to how well a teacher can accomplish objectives defined by someone else.

There seem to be two distinct methodological approaches to the measurement of teacher effectiveness: one is to look at what the teacher does, that is, to look at the process he uses; and the other is to look at what the teacher accomplishes or achieves--that is, at what his pupils learn, or, in a word, to look at the product.

It would seem obvious that if we look at the process--at how a teacher acts while he is teaching--we cannot hope to measure directly anything more than his skill as an instructor. The effects he is having are not visible.

In order to estimate a teacher's effectiveness indirectly by looking at the process he uses--that is, by observing the teacher as a basis for evaluating him--one must know more about the dynamics of teaching than research can tell him. He must know rather precisely what behaviors can be safely expected to have what effects on pupils; he must know and take into account the various characteristics of individual pupils which determine which behaviors are effective with which pupils; he must know and take into account the characteristics of the teacher, the content being taught, the objectives, and any other things which may affect pupil learning in the presence of any given teacher behavior. Not an easy way to go about the task! Perhaps it would be better to attack the problem head on by looking at the product--at how much the pupils are learning. Let's see what the research literature has to say about measuring teacher effectiveness in terms of pupil gains. But before we do so, I would like to digress a bit and talk about relationships between the two kinds of measures.

One of the disturbing things one finds as soon as he turns to the literature is that whatever the thing we assess when we look at process may be, it is something quite different from the thing we assess when we look at the product. The teacher who is rated highly effective (or skillful) by his immediate superiors

is not much like the one whose pupils seem to be learning the most according to gains on standardized tests.

Let me share with you the results of a study we did in New York City some years ago, which we found on checking are typical of other investigations of the same problem. What we did was to collect various kinds of information about teacher effectiveness in helping elementary-school pupils learn to read. We asked the person responsible for supervising each of 49 first-year teachers to estimate how that teacher would rank among typical first-year teachers he had known in ability to help pupils improve in basic skills. We also asked each teacher to estimate where she would rank herself in such a group (all but three of these teachers were females). We also administered a questionnaire to the pupils taught by each teacher which yielded (among other things) an index of how well the teacher was liked by her pupils. And, finally, we tested the pupils in the fall with the California Test of Mental Maturity and the California Reading Test, and tested them again in the follwoing spring with a different but equivalent form of the Reading Test. By analysis of covariance we then estimated the mean gain in reading test score for each teacher's pupils, making allowance for differences in pupil ability measured by the Test of Mental Maturity. We made all of our comparisons and estimated all of our correlations between teachers in the same shcools, thus removing all effects that differences in school populations might have on teacher effectiveness.

This is what we found when we intercorrelated these four measures of teacher impact. (Exhibit 2) We found clear evidence

EXHIBIT 2

INTERCORRELATIONS AMONG CRITERIA OF EFFECTIVENESS

CRITERION	PUP GR	SELF RTG	PUP RAP	SUP RTG
PUPIL GROWTH IN READING	.84	.41	.00	.13
TEACHER'S SELF RATING	.41	—	-.05	.12
PUPIL-TEACHER RAPPORT	.00	-.05	.89	.38
SUPERVISORS' RATINGS	.13	.12	.38	.89

that the four criteria were measuring two kinds of teacher effects which had little in common with each other.

Supervisors' opinions and pupils' opinions showed a significant intercorrelation of .38. Teachers' self-ratings and means of pupils correlated .41 with each other. But neither pair correlated with the other more than .13.

What this means is that the teachers rated high by supervisors and liked well by pupils were not the teachers whose pupils showed greatest gains and who judged themselves most effective. Teachers who looked most effective to supervisors were not actually the most effective in helping pupils learn to read. Process measures and product measures did not correlate with each other.

If we assume that the process measures used were valid measures of skill and that the product measure was a valid measure of effectiveness, we must conclude that teaching skill has little to do with teacher effectiveness! And let me mention that results of all similar studies we could find reached the same conclusion! Since different supervisors were in good agreement about which teachers looked best to them, we must assume that their judgments were objective and reliable. Only trouble is, they were basing their judgments on the wrong things. They agreed as to which teacher's pupils would learn most, but they seemed to be wrong--seemed to be basing their judgments on the wrong teacher behaviors.

If you were going to observe a teacher so you could judge how effective she would be, what would you look for? In the study I have been discussing we also sent trained observers into

the classroom to record the behaviors they saw without attempting to evaluate their possible effects on pupils. It was thus possible for us to compare the behaviors of those teachers judged poorest to get an idea of what kinds of behaviors the supervisors thought to be effective.

You might wonder why we did not ask the supervisors what they were looking for, what they based their judgments on. This has been done many times, and the results are consistent enough to be considered conclusive. Supervisors say they look for such things as the "ability to discipline, ability to teach, scholarship, and personality." Our actual observations indicated that a teacher was judged effective if her classroom was relatively quiet and orderly, and if there was little or no manifest hostility between teacher and pupil or pupil and pupil--which sounds like the "ability to discipline" to me. "Ability to teach," of course, is not yet definable in terms of observable classroom behavior. Nor is "scholarship" or "personality" so far as I know. Differences in "ability to discipline" (or whatever the supervisors saw) were great, and reliably measured. But they seemed to have little to do with the amount of learning that took place. Relatively high average pupil gains were just about as likely to take place in less orderly, friendly, and relaxed classrooms as in those where the pupils were busy, content, and task-oriented. It would seem, then, that there is only one way to attack the problem of measuring teacher effectiveness. The obvious way to do it would seem to be by examining the effects the teacher has on his pupils, and not worrying how.

You can scarcely believe all of the things wrong with this approach. I shall discuss only two. One of them has to do with reliability; the other with validity. Let me say something about the reliability problem first.

The general approach used in getting a measure of teacher effectiveness from pupil learning is to do what we did in the study I have already mentioned--use the adjusted mean gain of the teacher's pupils on some test or battery as the measure of effectiveness. There are refinements, of course, such as the use of covariance analysis to make allowances for differences between classes in such things as general intelligence; and more elaborate statistical refinements, which I do not want to get into today. We may assume that the statistical methodology is adequate.

We have assumed in all research up to now that teacher effectiveness is a relatively stable teacher trait; that the best teacher in a group this year will still be one of the best teachers in the group next year, even though he has a different class next year.

In a study of the teaching of reading in the first and second grade we did in New York City we obtained mean pupil gains in reading scores of classes taught by the same teachers in two successive years and correlated them. The magnitude of the correlation between two such sets of mean gains indicates to what extent the same teacher was ranked equally effective in two successive years. It is actually a kind of test-retest reliability of teacher effectiveness.

We used several different measures of pupil gains in the study so that there were nine reliability coefficients in all. The nine reliabilities ranged from a high of .53 to a low of -.08. The median was only .26, which was not even significantly different from zero!

Even the highest value of .53 is not very high, as reliability coefficients go--or should go; and the fact that four of the nine were significantly greater than zero is not very encouraging. That was less than half. We were taught in graduate school that a test used as a basis for decisions about individuals should have a reliability of .90 or so. If I apply the Spearman Brown formula to the highest value (.53) I find that it would take 8 years to develop a measure of teacher effectiveness with a reliability of .90 by this method. Are these results typical of those obtained in other, similar studies?

Rosenshine has reviewed all the other studies he could find which yielded stability coefficients on product measures of teacher effectiveness. Others' results tend to be consistent with ours, except that when the reliability was based on situations in which a teacher taught the same content to different students for 30 minutes (instead of a whole year) reliabilities ranged between .45 and .70. This is better, but not really as high as we would like to see it. We can summarize by saying on the basis of rather limited research that measures of teacher effectiveness based on pupil gains tend to be rather unstable.

So much for reliability. Can we get any feeling from the literature about the validity of mean pupil gain scores? We can if we look at the problem from the standpoint of content

validity; we must try to answer the question: do the tests used to measure gains provide adequate measures of the degree to which the teacher is achieving the goals of education?

I am afraid we must answer this question by saying, only if a rather narrow definition of the goals of education is used. If we are interested in only a single facet of achievement--such as learning to spell or to comprehend paragraphs--we can sometimes find a test which seems to be an adequate measure of that aspect of pupil achievement. But if we are measuring teacher effectiveness for evaluation purposes, as I assume most of you will be, we need to measure effectiveness in achieving most, or at least a good share, of the things teachers are supposed to do. If we include ability to help pupils develop attitudes and values or acquire inquiry skills (for example) as part of what an effective teacher does, it is quite clear that measures of pupil gains are and must for a long time remain lacking in content validity because of the lack of valid tests of these characteristics.

If a school system is willing to adopt achievement on some test, test battery, or combination of tests as sufficiently representative of the totality of its goals to approximate a measure of accomplishment of the total aims of the school, a valid and reliable measure of teacher effectiveness can in theory at least be obtained by measuring pupil gains for several years. But in terms of reasonably adequate sets of goals and periods of time, the early evidence is that no such measure is practicable at present.

As you must know, this position is somewhat out of fashion in these days of accountability, criterion-referenced tests, and performance contracting, when some school systems seem to be willing to settle for specified levels of gains on certain paper-and-pencil test items as criteria for assessing effectiveness not only of teachers but of the school itself. This represents an abdication of responsibility for many kinds of learning which I consider at least as important as those retained (if not more so). As a measurement specialist I am appalled. But that is not our topic today so I will dismiss the subject with the remark that if you accept this model for education the problem of measuring teacher effectiveness ceases to exist.

I would like briefly to mention another approach which has great appeal on the surface. I refer to what is usually called a "teaching test." In this approach the teacher is given a brief period of time--perhaps half an hour, perhaps two or three class periods, in which to teach a certain unit of content to a class which is tested before and after so a mean gain can be calculated.

Popham has done the most work with this approach. He has tried to validate it by comparing performances of trained teachers and various groups of persons with no professional training--college students, housewives, automobile mechanics. In no case has he been able to find any evidence that the trained teachers do any better as a group than the lay groups.

You may regard this as an indictment of teacher training as worthless. I maintain that the ability to cram content into

pupils' heads long enough so that they can score high on achievement tests is not what teacher education is trying to develop. Teachers who can do that are not the kind of teachers I evaluate highly. Scores on such teaching tests are worse than the pupil mean gain scores we have been taking about.

On the whole, I think we should give up the idea of measuring teacher effectiveness in terms of pupil gains on tests, attractive though the idea may seem on its surface. Let us take a look at process evaluation as an alternative approach.

Most efforts to measure effectiveness in process use rating scales of some sort. I don't want to get off on the question of what is wrong with ratings, except to mention one problem--the problem of determining what determines a particular rating. When we observe a teacher at work, all we can see is his skill. His effectiveness must be inferred, and such inferences (as we have seen) depends on an assumption that skillful teachers are effective. So they are, if they possess the right skills. To rate effectiveness, an observer must know what skills or behaviors make a teacher effective, and he must base his rating on them.

What is the present state of our knowledge of the nature of effective teacher behavior--of what you can expect to see in the classroom of an effective teacher? We have made some progress these last 20 years, but we still have a lot to learn.

The late A.S. Barr devoted his professional life to the search for behaviors more likely to be observed in classrooms of effective teachers than inefficient ones. The search was in vain.

In a classic review of all research done up to 1950 or so,

including Barr's, Morsh concluded that:

> No single, specific, observable teacher act has yet been found whose frequency or per cent of occurrence is invariably significantly correlated with student achievement.

Such specific, observable acts would, of course, be exactly what we need in order to assess teacher effectiveness in process--some concrete things on which we could base our evaluations. Apparently they do not exist.

Benefiting from Barr's experience, research into the nature of effective teacher behavior has taken a different tack during the last decade or two. No longer do we look for these universal behaviors. Harold Mitzel and I pointed out in 1962 what everyone else seemed to know already--that the effect of a particular behavior was specific to the pupil who was supposed to be affected, the teacher who was trying to affect him, and the situation in which it occurred. We also suggested that behaviors which contribute to teacher effectiveness would manifest themselves not as isolated events but as stable tendencies or patterns which have sometimes been referred to as elements of teaching style.

Beginning with the work of John Withall around 1948, research in the teaching process took a new direction. Observers still concentrated on recording what they saw happening in the classroom, as Barr and his associates had done, but instead of concentrating on individual behaviors they looked for dimensions or patterns manifested in various specific behaviors which

tended to occur together in the same classrooms. I am referring to such things as Withall's Index of Social-Emotional Climate or Anderson's Dominative and Integrative Contacts, or Flanders' ID ratio. Researchers like Flanders, Bellack, Smith, Hughes, Galloway, Gallagher and Aschner, Spaulding, and others, who have in twenty short years so greatly increased our understanding of classroom behavior. We still do not know nearly as much as we need to know about what constitutes an effective teaching style, but we can describe the style of any given teacher much more objectively and accurately than we ever could before.

The more dimensions of teacher behavior we can measure and correlate with pupil gains, the more likely we are able to find one or more that correlates with teacher effectiveness. And when we do find one that does correlate, we already have, not only a clear operational definition of the dimension of behavior but also a device for measuring it.

Have we located any such dimensions? Or, to put it differently, have we identified any observable characteristics of teachers on which we can defensible base judgments of their effectiveness? Knowledge of such characteristics might tell us what to look for and so improve our ability to evaluate teachers by observing them at work.

In discussing this problem I shall draw heavily on excellent and recent reviews of empirical studies already done by Rosenshine and Furst and by Haroutoonian, as well as the earlier reviews done by Morsh and Barr some years ago.

Before I start let me emphasize that none of the

relationships I shall describe between observable characteristics and pupil achievement gains can be said to be firmly established; each one included in the list has, however, been found to show some relationship to pupil learning in more than one independent investigation. The best way to think of them is as promising leads. In order to put them in better perspective, it might be useful first to list some observable characteristics or behavior patterns which researchers have tried to measure and relate to pupil gains without succeeding. (Exhibit 3)

Here are six kinds of measures that have been tried with negative or inconsistent results. And yet there is a viable reason for expecting that each one would be related to effectiveness.

Whatever one's theory of the nature of teacher effectiveness may be, one would expect it to grow with experience: but the research evidence does not indicate that pupils learn more from experienced teachers than from inexperienced ones. And it would seem even more reasonable to think that pupils who had less contact with a teacher would learn less from him than ones who had more, but we have not been able to verify this either. Even if one takes the position that there is no such thing as a science or art of teaching, one would expect that an ignorant teacher's pupils would tend to learn less than pupils with well-informed teachers. Perhaps so, but no clear evidence has been found that they do.

Psychologists like B.F. Skinner claim to have shown a close relationship between reinforcement and learning in rats, cats, and pigeons, and have not hesitated to recommend that teachers reinforce pupils by approving and praising correct responses.

SOME TEACHER CHARACTERISTICS NOT FOUND TO RELATE TO TEACHER EFFECTIVENESS (UP TO NOW)

1. EXPERIENCE

2. ATTENDANCE

3. KNOWLEDGE OF SUBJECT

4. PRAISE-APPROVAL-WARMTH

5. I/D RATIO

6. PUPIL PARTICIPATION

EXHIBIT 3

Teachers who use a lot of praise and approval do not seem to get any better results, however, than teachers who do not. Psychologists like Rogers and Flanders have strong theoretical reasons for recommending indirect teaching, yet the I/D ratio, designed to measure this very characteristic, has not been found to predict pupil learning either.

All of us remember the Deweyism, "We learn by doing." Attempts to relate the amount of pupil activity (doing) in the classroom to gains in knowledge (learning) have not been successful.

If you find these results discouraging, you can imagine how the researchers who obtained them must have felt. All of these variables seem so obviously related to pupil learning as hardly to be in need of empirical demonstration. None of them has paid off. I am not, let me assure you, suggesting that we conclude that none of them are related to teacher effectiveness. What I am suggesting is that the variables named must not be as easy to identify or recognize as you might think--or else we need to re-examine these propositions.

Let us turn now to some measures which do show signs of being related to pupil learning. Some of them look very much like the ones we have been discussing, and some do not. (Exhibit 4)

If these results are taken as a guide, we might describe the behavior of the teacher whose pupils are likely to learn more than those in the average class as follows:

1. He varies the level of tasks assigned to pupils, the methods and materials he uses, and his strategy in presenting them. Accordingly, he asks more questions at higher cognitive

SOME TEACHER CHARACTERISTICS

WHICH HAVE BEEN FOUND (IN TWO OR MORE STUDIES) TO BE CORRELATED WITH TEACHER EFFECTIVENESS

Characteristic	Direction
1. VARIABILITY	HIGH
2. COGNITIVE LEVEL OF QUESTIONS	HIGH
3. PERCEIVED DIFFICULTY OF QUESTIONS	HIGH
4. ELABORATING (PROBING) QUESTIONS	FREQUENT
5. CRITICISM OF PUPIL COMMENTS	INFREQUENT
6. ACCEPTANCE/USE OF PUPIL COMMENTS	FREQUENT
7. CLARITY/ORGANIZATION OF PRESENTATION	HIGH
8. STRUCTURING STATEMENTS	FREQUENT
9. OBJECTIVES-RELATED STATEMENTS	FREQUENT
10. MANNER	TASK-ORIENTED BUSINESSLIKE
11. ENTHUSIASM	HIGH

EXHIBIT 4

levels than the typical teacher and his pupils perceive the tasks assigned them as difficult. He tends to ask his pupils to elaborate their own comments or those of other students. He tends to be sparing in his criticism of pupils' responses, but is more likely to accept and use them.

His own presentation of material is perceived by observers as clear and well organized, containing by actual count a relatively large number of structuring statements and statements whose content bears directly on specific objectives of the lesson. His behavior is described by raters as task-oriented or business-like and also as characterized by enthusiasm.

I can almost hear some of you saying to yourself, "I knew that. Who needs all of this research to tell us that an effective teacher does those things?" Let me repeat an important point. Of course we knew these things. We also "knew" that an effective teacher was one who had a substantial amount of experience, knew his subject well, developed content by interacting with his students rather than lecturing to them, and commended or praised his pupils when they gave acceptable responses. But for some strange reason, the research only supports the former characteristics, not the latter.

Could it be that the research is trying to tell us something? Is it trying to help us separate the things we know that are <u>true</u> from the things we "know" that are <u>not</u> <u>true</u>? Would we be wise to try in our evaluations of teachers to ignore the second kind of beliefs and look harder for the first?

One quality I seem to see in the verified list is a kind of complexity or sublety that is absent in the one not verified by

the research. Simple minded ideas like reinforcement, indirect teaching, learning by doing, experience makes the best teacher, the more practice the more learning, or the more the teacher knows the better he is, which are reflected in this list, just plain don't stand up.

This is a pity, because it is so much easier to comprehend these simple ideas, to communicate them to others, to recognize them when we are asked to observe a teacher and evaluate his performance, than the more complex characteristics which do seem to be related to teacher competence.

I envision two developments related to the measurement of teacher effectiveness which will be beneficial. I see ratings becoming more valid as we gain greater insight into the nature of the effective performance we are looking for from research along the lines reported in the last two exhibits.

The other is the substitution of more objective instruments for rating scales; instruments based on, and perhaps adapted from, the instruments used in the research studies. Use of observation schedules will provide the teacher evaluator with more accurate, relevant, and detailed information than we can get from any rating device.

I think, however, that we will soon approach a ceiling or plateau beyond which we cannot rise without a change in strategy. In order to make clear what I mean, let me show you a model of the dynamics of teaching skill. (Exhibit 5)

A teacher in a classroom trying to implement a plan must do at least three things all at once: he must maintain the learning

environment, he must manage learning activities, and he must involve individual pupils. The skilled teacher is something like a juggler or plate-spinner; he keeps all three of these things going with an occasional deft touch only--a touch so deft, sometimes, that the observer has difficulty seeing how it is done.

He does them all at once, of course--but it is convenient to think of them as hierarchically related. Let us imagine that in each cycle the teacher first checks the climate in the classroom and adjusts it if necessary. Then if climate is OK, he checks the status of the planned activity supposed to be underway. If that is not progressing according to plan he makes whatever adjustment may be necessary. If the plan is progressing on schedule, he checks the involvement of each and every pupil, and adjusts that. I suspect that the whole business never runs quite right, that there is always room for improvement somewhere.

Unless the observer charged with evaluating the teacher knows something about the model, he is not likely to make much sense out of what he sees. I suggest that observers, like the raters I mentioned earlier, respond mainly to the climate. Environmental maintenance behaviors may be consistent enough across time to be seen often enough so the observer spots them. Managing and involving behaviors may vary so much that it is difficult for a visitor who may not know exactly what the pupils are supposed to be involved in, or in what idiosyncratic ways they react to the activities, to detect them when they occur. Experience indicates, however, that the quality of the learning environment is accessible to the most casual observer.

There seems no reason to doubt that the learning environment a teacher provides has quite a bit to do with his effectiveness in general. Children are curious, lively, interested beings: put them into a situation where there are interesting materials and where the environment favors learning and they will learn on their own, and show measurable progress toward some goal.

So if we look at a teacher and decide he looks good, chances are we are reacting to the climate we detect in his classroom. If we see his pupils busily and contentedly working at their seats, participating in an orderly but lively discussion, etc., we conclude that they are probably learning and rate the teacher as effective. We are probably, but not necessarily, right. It depends on the extent to which we respond to the simpler, more obvious characteristics--indirectness, pupil participation, warmth and lack of hostility that as we have seen, have relatively little to do with pupil learning.

Hopefully, by looking for some of the less obvious things--the cognitive level of questions asked, probing questions, clarity and structure--we can sensitize ourselves to the more important dimensions of classroom climate and improve the validity of our ratings somewhat.

If, however, you accept my model of teaching skill, or something like it, you will agree, I think, that until we devise some economic and practicable scheme for assessing a teacher's skill in conducting planned learning experiences and seeing that individual pupils are involved in them, I doubt very much whether we will be able to get at the important differences in teaching

skill. No matter [illegible] successful an elementary teacher may be in providing a classroom environment which facilitates pupil learning, she will never get her pupils to learn anywhere near their capacity unless she can conduct them through experiences designed to produce optimal learning of reading and arithmetic skills, for instance, so that they can exploit that environment.

EXHIBIT 5