PROBLEMS OF TEACHER MEASUREMENT

B. F. PITTENGER
University of Texas

Teacher measurement, in the sense of estimating the efficiency of teachers individually, is not a new and theoretical proposal, but an existing supervisory responsibility. This statement needs no demonstration. But some discussion may be necessary of the further statement, that, in so far as teacher measurement is at all reliable and consistent it must be controlled by some sort of schedule or list of items representing qualities of teaching merit. Such a schedule may never have been set down on paper; indeed, its possessor may hardly be aware of its existence. But it is humanly impossible for a supervisor to report from year to year upon the efficiency of the various members of his teaching staff, accepting the responsibility for their retention, advancement, demotion, or dismissal, and particularly for their assistance and improvement while in service, without having formulated, clearly or crudely, somewhere in his mind, a statement of the factors which in his opinion constitute efficient teaching.

The current efforts of experimentallists in the field of teacher measurement are only attempts to extract from the consciousness of principals and supervisors these personal criteria of good teaching, and to assemble and condense them into a single objective schedule, thoroughly tested, by means of which every judge of teaching may make his estimates more accurate, and more consistent with those of other judges. There is nothing new about the entire movement except the attempt to objectify what already exists subjectively, and to unify and render universal what is now the scattered property of many men.

There are those who believe that the movement toward teacher measurement is a monstrous innovation, which threatens the holiest traditions of the educational profession by putting a premium upon mechanical methodology. It is not an innovation. But the phrase "teacher-measurement," itself, no doubt, is in part responsible for this misunderstanding, as it suggests a mathematical exactness of procedure which is clearly impossible in this field. Teacher measurement will probably never become more than a carefully controlled process of *estimating* a teacher's individual efficiency. The phrase is, however, sufficiently con-

venient and euphonious, and has now been used widely enough, to warrant its continuation.

To students who are familiar with the reports of current studies in this field there occur objections to present procedure of a more fundamental character. However sympathetic one may be with the general plan of devising schedules for teacher measurement, it is difficult to justify many of the methods by which these investigators have attacked the problem. For example, all of them appear to have set up as their goal the construction of a schedule which can be applied to any teacher, whether in the elementary or high school, and irrespective of the grade or subject in which his teaching is being done. "Teaching is teaching," is the evident assumption, "and the same wherever found." But it may reasonably be maintained that different qualities and methods, at least in part, are requisite for a teacher's success in the primary as contrasted with the grammar grades, and that the criteria of good teaching are not entirely the same in the different departments of the high school and of yet higher institutions. In so far as the criteria of good teaching *are* the same in these very diverse situations, it seems probable that the comparative importance to be attached to each must differ. If these contentions are correct, the problem is to devise, not a single blanket schedule for universal application, but a series of independent schedules for teachers of different grades and subjects in the elementary and high schools. If, on the other hand, these statements are not true, and if a single schedule can actually serve in place of many, the surest way to discover that fact is to begin by constructing schedules for specific grades and subjects. Only if the different specific schedules eventually prove to be alike, can a single blanket schedule wisely be substituted for them.

The second major criticism to be urged upon existing studies of this sort is that there is an unnecessary and wholly unjustifiable overlapping among the various items or rubrics comprising the proposed scales. Logically, there are three possible planes upon which an estimate of teaching efficiency may be constructed: (1) the plane of results or of pupil achievement; (2) the plane of the teaching and learning process; and (3) the plane of the teacher's equipment for teaching, both native and acquired. This kind of analysis appears not to have been made by the

investigators whose work has been reported. Certain items in their schedules have been drawn from the plane of results (for example, "the growth of pupils in subject-matter"),[1] others have been taken from the plane of the schoolroom process (such as "skill in stimulating thought"), and still others have come from the plane of the teacher's equipment ("academic" and "professional preparation"). Clearly, in so far as elements in the schoolroom process are real factors in teaching efficiency they are measured once in the measurement of results. To measure them again, independently, is to measure them twice. Similarly, academic and professional preparation, in so far as they have actually affected teaching, and hence have become real factors in efficiency, are measured once when results are measured, are measured again when the teaching process is measured, and are measured yet a third time when measured by themselves. In precisely the same way, results are measured indirectly in the measurement of the teaching process, and are again measured indirectly when one measures the teacher's equipment for his work. Assuming that the various items representing each plane of measurement have been selected at random, but that none of the planes is represented completely by the items chosen, we have no *a priori* means of determining which of the respective items have, in reality, been measured three times, which have been measured twice, and which have been measured only once.

Not only has there been an unrecognized overlapping of the levels of measurement from which the individual items of the schedules have been taken, but there has been equally unrecognized overlapping among the items representing each level. To take a single illustration: In Boyce's scale, among the items representing the plane which we have described as "the teacher's equipment," we find the rubrics "health" and "voice." Ordinary observation alone is sufficient to convince one that health is frequently a large factor in determining both the quality and the control of a teacher's voice. This one illustration could be multiplied many times. Indeed, there is hardly an item in this or in the other scales which does not clearly overlap upon one or more of the other items making up the schedule.

[1] The examples given in parentheses are taken from the tentative schedule proposed by BOYCE, in Part II of the Fourteenth Year Book of the National Society for the Study of Education, page 45. Illustrations of the same sort of combination might be drawn from other published schedules.

It would, of course, be a hopeless task to attempt the construction of a schedule in which all of the items would be mutually exclusive. But it is highly desirable, in fact it is necessary, to reduce this overlapping to a minimum, and to utilize such methods as are available to estimate or to eliminate the effects of unavoidable duplication. Investigators in this field have not only permitted the grossest overlapping to occur among the component items of their scales, but they have failed to make use of available methods for eliminating its disturbing consequences.[2]

In the third place, students of teacher measurement appear to have erred in that they have attempted too much. The writer is strongly of the opinion that, for the present at least, efforts to construct a schedule for teacher measurement should be confined to a single one of the three planes which have been enumerated. Doubtless in the end we shall want to know as much as possible about all three; and to combine in our final estimate of a teacher's merit all attainable facts as to her equipment, her classroom procedure, and the results which she achieves. But at present we should do wisely to project our investigations upon one plane at a time, and to make each of these investigations as thorough as it is possible to make it. Later, when we know the nature and comparative value of the various items necessary to adequate judgment upon all planes, there will be time and opportunity for putting together the different schedules into one.

Let us endeavor to make this matter concrete. Let us suppose that each of the planes referred to contains ten elements of efficiency; in other words, that there are ten elements in the teacher's inherited and acquired equipment, ten elements in an effective teaching process, and ten requisite elements of results. Is it not clear that a schedule made up of ten elements comprising, for example, the plane of the teaching process, would be a more permanent achievement than a schedule made up of ten items distributed over all three planes? A schedule made up of items confined to one plane need not be disturbed by advances subsequently made upon the other planes, but a schedule whose

[2] Attention may here be called to the mathematical formula known as the "regression equation" which has been used by T. L Kelley, C. T. Gray, and others, to eliminate the complications arising from the sort of difficulty that we have described.

items are distributed over all three planes must be completely revised with every discovery made upon any plane. The great advantage of confining endeavor to a single level at one time is that results secured upon that level can stand undisturbed by contemporary or future achievements made upon different levels. It is customary and wise to lay the foundation securely before rearing the house, and not to try to construct both the foundation and the superstructure at the same time.

While other cogent reasons might be advanced for urging the concentration of an investigator's energies along a single plane, the one that has been given is sufficient. Without further argument of the point, therefore, let us try to determine which of the three levels offers the best prospect of useful results.

The plane of results (in the sense of changes wrought in pupils) would be the ideal plane upon which to build an estimate of a teacher's individual efficiency, if it were possible (1) to measure all of the results of teaching, and (2) to pick out from the body of measured results any single teacher's contribution. At present these desiderata are impossible to attain. Only the more mechanical products of a teacher's efforts can now be subjected to accurate appraisal, and means exist for the measurement of only a small portion of these. This fact is no indictment of present efforts in the field of educational measurements. Let us continue to measure such results of teaching as we can, and let us by all means extend our facilities for this purpose; but let us not make the mistake of assuming that the results that we can measure are the only results of teaching, or even that they are the most important part.

Moreover, we must keep constantly in mind the fact that the results which pupils achieve in any given subject are by no means the product of the labor of any single teacher. Earlier teachers, other contemporary teachers, and the environment external to the school, are all factors in determining pupil efficiency in any school subject. It has been urged that the influence of these complicating factors can be materially reduced by measuring only the change in pupil achievement which takes place under the guidance of a single teacher. But it must be remembered that this process only reduces these complications; it does not and cannot eliminate them.

It seems clear, therefore, that we are not now in a position to base an estimate of a teacher's individual efficiency upon the measured power of accomplishment shown by her pupils, nor even upon the measured change in the power of accomplishment which takes place under her tuition. We have not the facilities for measuring all of the results of teaching, nor for determining for what part of the measurable results any single teacher is responsible. Under such circumstances, we must turn to the other levels which have been described in search of more encouraging possibilities.

Of the remaining levels, that of the teaching process seems to offer greater promise than that of the teacher's equipment for teaching. The test of a teacher's efficiency is not so much what she *can do* as what she *does*. That teacher is inefficient who is doing inferior work, no matter what the standard of work she may be able to maintain; and that teacher is efficient, though certainly not the most efficient, who is doing good work, irrespective of her ability to better it. But when a teacher's present efficiency has once been ascertained, facts relative to her latent abilities become of great significance, as the means for guiding her progress upward to higher planes of usefulness. The securing of these facts, however, constitutes a separate problem.

Our discussion has now brought us to the level that we have called the plane of the "classroom process." It is upon this plane that the writer recommends immediate and concentrated investigation, This level is superior to the level of results in that it is here possible to differentiate more clearly the activities of a single teacher, and it is superior to the plane of equipment in that it represents actual and not potential efficiency. It is also, in all probability, more readily measurable than the other planes. However, measurement upon this level is not without its serious difficulties.

One of these difficulties is that of defining accurately the meaning of the classroom process. Does it include teacher-activities, or pupil-activities, or both? There is good reason for urging that it be confined to pupil-activities. Pupil-activities alone are responsible for the growth of pupils, and it is for stimulating the growth of pupils that schools and teachers exist. On the other hand, pupil-activities are to no small degree a joint product, due to influences flowing from all the teachers in the school, and

from agencies outside the school. No one teacher can be given the entire blame or credit for the doings of the pupils in her classroom. Only the teacher's own activities represent herself.

The common-sense answer to our question seems to be that the "classroom process" should be regarded as including the activities of both teachers and pupils. This is the meaning adopted by the practical supervisor, who, when visiting a classroom for the purpose of judging the quality of instruction, looks at both teacher and pupils for indications of what is going on. Why should the schedule-maker do less than this? It is the purpose of a schedule, not to rob a supervisor of any of his sources of information, but to guide him in the use of these sources to the end that he may draw from each one all that it has to give.

If the reader accepts the propositions that have been advanced thus far, he should be ready to accept the suggestions which follow for the guidance of future research in the field of teacher measurement. Let the investigator select a single school grade or subject, and confine himself for the time to the elaboration of a schedule for teacher measurement in that alone. Let him also limit his efforts to a single plane of measurement, for the time at least; preferably to the plane of the classroom process. Let him not attempt the visionary task of developing a schedule for measuring all teachers, in all grades and subjects, and upon all planes.

There is one apparent objection to these proposals that should be mentioned. It would be impossible, under this policy, to reduce one's estimate of a teacher's efficiency to a single statement, such as a mathematical per cent. The ratings given a teacher in native ability and preparation, in classroom efficiency, and in results as manifested by the achievements of her pupils in standard tests, would have to be kept strictly independent of each other for a long time to come. Not knowing their relative importance, it would be impossible to transmute these different statements of efficiency into one. But the writer is unable to see that this fact forms a valid criticism. Hasty reduction of these various statements to one is to secure a final statement that is false. Furthermore, what is the practical utility of a single mark for representing a teacher's efficiency? Would any sane supervisor recommend the discontinuation of teacher's services simply because she fell below a stipulated percentage mark in

terms of any schedule? Would any responsible board accept such a recommendation if it were made?

A supervisor's judgment of the final merit of each of his teachers must be formed in the future as it has been formed in the past. Assembling in his mind all of the obtainable facts with regard to a teacher's work, he must form his opinion as to her needs and her deserts in the light of these facts plus his knowledge of the needs of his particular school. The promotion, discharge, or constructive criticism of teachers cannot be reduced to mathematical formulae. The proper function of a scorecard for teacher measurement is not to substitute such a formula for a supervisor's personal judgment, but to aid him in discovering and assembling all the data upon which intelligent judgment should be based.