

PPA DELTA PI RESEARCH PUBLICATIONS



THE MEASUREMENT OF TEACHING EFFICIENCY

The Measurement of Teaching Efficiency

BY

WILLIAM H. LANCELOT,
ARVIL S. BARR, THEODORE L. TORGERSON,
CARL E. JOHNSON, VERGIL E. LYON,
ANTHONY C. WALVOORD, AND
GILBERT LEE BETTS

EDITED BY

HELEN M. WALKER, PH.D.
SPECIALIST IN EDITORIAL CHARGE OF THIS VOLUME

NEW YORK
THE MACMILLAN COMPANY
1935

FOREWORD

WITH the publication of this volume Kappa Delta Pi begins its Research Publications. As an honor society in Education Kappa Delta Pi seeks to foster scholarship in the broad field of professional education of teachers. To this end the Society created its Lectureship Series in 1928. The present series of monographs was authorized at the Convocation of the Society held in Atlantic City in February, 1930. At that time the sum of one thousand dollars was set aside as an award for the best report of research during the next bien-nium on a subject to be assigned by the Executive Council and as recommended by its Committee on the Research Award.

By action of the same Convocation the Executive Council was further authorized to publish in a research or monograph series the study receiving the award. If no study was deemed sufficiently meritorious to receive the award the Executive Council was authorized to publish summaries, abstracts, or abridgments of those studies adjudged sufficiently worthy of publication in this form. Authority was also given the Executive Council to waive the bestowal of the award if none of the studies submitted seemed to be a wholly adequate contribution to the assigned field of inquiry.

The Executive Council was also empowered to administer the Research Award according to such plans as would insure competent adjudication of the merits of each study submitted. Thus empowered the Council turned to the Society's Laureate Chapter and requested members of this chapter to select the study which they deemed of greatest value as a

contribution to the assigned problem. Professor Edward L. Thorndike and Professor William Heard Kilpatrick of Teachers College, Columbia University, and Professor Truman Lee Kelley of the Graduate School of Education, Harvard University, consented to serve on the reviewing committee. In addition to these judges, Professor Helen M. Walker of Teachers College, Columbia University, was chosen as special reader of the statistical content of the competing studies and editor of the proposed volume.

In the circular announcing the award the Executive Council defined certain conditions that should govern the participants and those making the award. Among these conditions the following are of special interest:

1. No single method of investigation is specified. The research may be experimental, statistical, or philosophical in character. The pertinency of the method to the phase of the problem studied will be considered in judging the reports.
2. The basic criteria in judging the worth of a report will be its contribution toward the solution of the problem, the validity of the techniques employed, and the organization and literary merits of the report.
3. The reports submitted in competition will be judged first by the Executive Council of Kappa Delta Pi. The three, four, or five reports (depending upon the total number received) adjudged to be best will then be submitted to a committee of the Laureate Chapter of Kappa Delta Pi and consulting specialists, who will select the winning report.
4. The report which receives the award will become the property of the Society and will be published by the Society in a monograph series complementing the present Kappa Delta Pi Lectureship Series. The Society may

publish in the monograph series or, in abridged form, in the *Kadelpian Review*, meritorious reports submitted in the competition which are not awarded a prize. Such obligation would be without expense to the author.

5. The Society reserves the right to reject all reports if, in its judgment, none is worthy of an award, and to divide the award between two contestants in the event that two reports have apparently equal merit.
6. The competition is open to anyone, anywhere — but all reports submitted must be in the English language.

The problem selected for the first competition was The Measurement of Efficiency in Teaching. Twenty-three competing studies were received. The reviewing committee noted several interesting manuscripts but found none of such preeminent value as to merit the award. Three of the studies which attacked the problem in three different ways were, however, deemed sufficiently significant to be recommended for publication together in a single volume. The authors were requested to make specific revisions: a reduction of the length of each report, amplification of those sections where the argument seemed inconclusive, elimination of certain doubtful techniques or unsupported deductions, and clarification of the presentation. Professor Walker, in addition to her critical reading of the statistical material, arranged the manuscripts in logical sequence. Neither the committee nor the editors have altered the essential plan of any of the reports.

The Measurement of Teaching Efficiency is one of the most challenging problems in the field of educational research. Although in the present volume the subject is treated experimentally and statistically, it admits, no less, of philosophical and sociological treatment. Even if adequate quantitative

measures elude the experimentalist, many important problems in Education might be solved if teachers could be classified into qualitative categories on the basis of teaching efficiency. Such quantitative and qualitative approaches, however, need the guidance of an understanding of what good teaching involves and criteria which will satisfy valid measures of it.

The studies in the present volume chart three approaches to a measurement of teaching efficiency. The first study reports an effort to determine certain permanent outcomes of the instruction given in a Department of Mathematics to classes of engineering students during the nine-year period, 1920-1928. In the second study the validity of certain instruments for the measuring of teaching ability is tested. The third study reports attempts to discover and to measure a complex trait of teachers and to discover the relationship between pupil achievement and this trait in teachers. The reports are detailed and the methods of investigation are lucidly described.

The Society of Kappa Delta Pi as a whole and its Executive Council in particular hope that the present volume and the evolving publications will prove helpful to educators in their efforts to solve vitally important problems in educational research. On behalf of Kappa Delta Pi and its Executive Council the editor of the publications thanks the authors, the reviewing committee, and Dr. Walker for their painstaking co-operation toward the launching of the Research Publications at this time.

ALFRED L. HALL-QUEST

GENERAL EDITOR OF THE

KAPPA DELTA PI RESEARCH PUBLICATIONS

PREFACE

THIS volume is significant not so much for the finality of its findings as for the way in which it illuminates the present status of the problem of measuring teaching ability. Because the authors of these three studies state repeatedly, with admirable frankness, the limitations of their data, and because they make it clear that their studies do not arrive at an ultimate solution of how to select the most successful teachers, the reader should not fail to recognize the real contribution which they have made. The reader who hopes to find here a blueprint giving him a short and easy way to judge the efficacy of teaching will be disillusioned. The more thoughtful reader who is willing to try to understand the all but insurmountable difficulties of the problem will find in these studies relationships worthy of his careful attention, as well as new and stimulating methods of attack.

The problem set by Kappa Delta Pi for this competition is one of most strategic importance for the advancement of education. When it is possible to measure teaching success, we shall immediately possess a valid criterion by which to evaluate all those matters which condition the nature of instruction as affected by school and class management. On the one hand are matters as widely varied as administrative procedures, programs of teacher training, curricula of teachers colleges, admission of applicants to teachers colleges and the relationship between the ability to teach and age and sex; on the other hand are the various personality traits, intelligence, and academic training. In fact almost every

practice or policy involved in running any part of the educational system would be open to review and reappraisal in the light of information about teaching success. Without such a measure none of the matters referred to can be studied with respect to their central function, the promotion of better teaching. Besides the practical need for such a measure whereby the wisdom of our day-by-day practices might be judged, there is a crucial need for it in a large variety of research studies whose conclusions are jeopardized when teaching ability is left an uncontrolled variable.

Several difficulties stand in the way of a completely satisfactory solution for this problem, and we may well examine a few of these and ask how the studies here reported have dealt with them.

Reliable identification of a trait is essential to its measurement. Verbal definition of the trait can often be dispensed with, but any trait which eludes identification of some sort, either by direct recognition or by recognition of its consequences or concomitants, can scarcely be measured. Unfortunately for the purposes of this research, educators do not agree very well as to who is a good teacher or what are the concrete manifestations of teaching ability. This would not be an insuperable difficulty if agreement could be reached concerning the desired outcomes of teaching, because then teaching could be measured not directly through measurement of the teacher but indirectly through measurement of pupil change. While the diversity of prevailing philosophies of education thus appears to be a bar to the construction of any universal measure of teaching ability, it may be possible to construct a measure useful within a specific frame of reference.

The lack of an adequate, concrete, objective, universal criterion for teaching ability is thus the primary source of

trouble for all who would measure teaching. One typical method of attack used in rating scales is to compile a list of broad general traits supposedly desirable for teachers, with respect to which the rater passes judgment on each teacher. This amounts to an arbitrary definition of good teaching, which is subjective and usually vague, but it does not necessarily lead to an identification of it. Only if the traits themselves can be reliably identified can their possessor be identified as a "good teacher" according to the definition laid down in the scale. Even when the scale is made quite specific, relating not to general traits but to concrete procedure, the fundamental difficulty remains, that there is no external and generally accepted criterion against which the scale can be validated to establish the significance of its items.

The experimenter seems to be on safer ground when he uses some form of pupil change as the main criterion for teaching success. If schools exist for the sake of children, then teaching should be judged by its effects upon children. But at once we are forced to decide which changes shall be isolated for study. Though readily available, easily administered, and in general developed to a higher state of technical excellence than tests of personality and attitude, nevertheless tests of subject-matter achievement are only partly satisfactory. Philosophically their use as the sole criterion of pupil change may be criticized as likely to result in the selection of too many good drill masters who have no other desirable qualifications, and in the exclusion of too many stimulating personalities who are not drill masters. Experimentally their usefulness is greatly limited by the suggestion, found in almost all empirical studies of this matter, that apparently a child's subject-matter achievement is more closely related to his own ability and previous record than to any help a teacher can bring to him. There-

fore pupil change on achievement tests usually shows a low correlation with any direct measure of the teacher. This is disappointing to the experimenter but worth much pondering by the educator. The measurement of pupil change in most non-scholastic traits, however, is as yet not very successful, the problems of identification and of discovery of a criterion in such traits being quite similar to those discussed previously in regard to the direct observation of "teaching ability."

The three studies in this volume make three very different and significant approaches to this problem of finding a criterion. The first study employs two novel criteria, entirely objective and definite, based on a measure of pupil change which involves both academic success and attitude without the use of pencil and paper tests of either, and these are related to long-range rather than to immediate outcomes. Though admittedly not symptomatic of all aspects of teaching and not feasible in all situations, these are criteria which most readers will consider reasonable, and their discovery is a contribution which promises to be of genuine importance. The second study accepts and utilizes the more common criteria of pupil gain on standardized tests, of teacher ratings, and of scores on tests of teacher traits, examining nineteen such measures already widely known, and it studies the relationship of each with the others. This study offers no new criteria but furnishes valuable data for the appraisal of those already familiar. The third study uses experience as criterion and validates a new instrument by means of its ability to differentiate two criterion groups of teachers, one a group of experienced and reputedly superior teachers and the other a group of novice teachers. There is, of course, no reason why the novice group should not have contained some teachers with great potentialities, and their presence would tend to

obscure differences which the instrument might otherwise detect in groups that did not overlap. This method, so often used in other situations, has been strangely neglected in studies of teaching ability, and is capable of much more extensive application there.

Another discouraging aspect of the attempt to find a measure which will show high correlation with pupil change is the apparent tendency of pupil change to be conditioned by a very large number of other variables. These may include such governing factors as intelligence, the pupil's own habits of study, interest, and physical condition. Pupil change may also be affected by factors associated with the teacher, his personality, voice, dress, clarity of thought and expression, sense of humor, and so on. In addition to the foregoing are a large number of factors such as the size of the class, the physical condition of the building, and so on and on. Let us say there are fifty such factors, though of course no one knows how many there are. Unless these fifty are themselves closely related, then pupil change must inevitably show a low relationship to most of them. It is therefore not to be held against a research worker that he fails to find a high correlation between pupil change and other measures studied. Very possibly no such high relationship exists except with measures connected with the pupil's own capacity and background. If, however, each of the fifty hypothetical factors could be studied to find the optimum conditions for pupil change of the sort desired, then there might be a very great difference between pupil change when all the conditions were optimum or nearly so, and pupil change when all the conditions were at or near the nadir, and this difference might have profound educational value.

Two different types of research, both carried out on a large scale, seem to the editor to be needed now. One is a survey

on a grand scale, in which a large number of teachers in fairly homogeneous school environments are given a great variety of tests of personality, intelligence, information, attitude, etc., and their pupils are given a large battery of tests including intelligence, achievement, personality, and attitude tests, these being administered both at the beginning and at the end of the period of instruction. All possible relationships are then investigated between teacher traits and pupil change, intelligence and initial test score of pupil being held constant. To carry this out on a really large scale, paying teachers to take the large number of tests required of them, would call for a very considerable subvention.

For the second research two criterion groups would be needed, one composed of very superior teachers, the other of teachers so poor that they are likely to be dismissed at the end of the year. If the conditions necessary to employ Dr. Lancelot's criteria are present, the group may be selected on that basis. Otherwise, those rare outstanding teachers who would rank as superior under any definition of good teaching must be sought. Since it is not necessary to assemble them in any one location, they can be chosen at one time wherever found. The other group should be made up of teachers so poor that almost any supervisor would consider them failures. A corps of trained observers may be sent into their classrooms, employing the new techniques of controlled observation to produce pictures of the good and the poor teacher in action, to see what it is which differentiates them reliably. This would take a long time and require the work of many observers, trained to a high degree of skill in observational techniques. At the present time in all probability there are scarcely enough trained observers in the country to carry it through, but the method seems promising.

Having assumed responsibility only after the studies were

submitted in completed form, the editor feels constrained to disclaim all credit for the fundamental plan of these studies or for the methods of analysis, and she reserves the right to express a minority opinion at certain points. In cases where a difference of opinion between author and editor could not be worked out in correspondence, the author's version has been allowed to stand and the editor's reservations consigned to a footnote.

That the present volume cannot offer a neat and final program for an objective method of selecting superior teachers should surprise only those persons who do not understand the nature of such research work. That its studies carry us several steps further toward the solution of an exceedingly important and exceedingly difficult problem, will be gratefully acknowledged by the experienced research worker.

HELEN M. WALKER

TABLE OF CONTENTS

I

A STUDY OF TEACHING EFFICIENCY AS INDICATED BY CERTAIN PERMANENT OUTCOMES

	PAGE
I. SCOPE AND NATURE OF THE INVESTIGATION . . .	3
Standards used in measuring the results of teaching. Differences in remote outcomes as measure of teaching efficiency.	
II. THE METHOD OF INVESTIGATION	7
Equating test groups as to general ability. Unequal distribution of teachers in subsequent courses. Making up the test groups. Measures studied.	
III. RESULTS SECURED IN TEACHING MATHEMATICS 1 .	12
Results of teaching superior students. Treatment of differences. Problem of combining the two standards. Arbitrary basis of judging outcomes. Results of teaching average students.	
IV. RESULTS SECURED IN TEACHING MATHEMATICS 2 .	24
Results of teaching superior students. Results of teach- ing average students. Special adaptations of teachers.	
V. RESULTS SECURED IN TEACHING MATHEMATICS 3 .	32
VI. RESULTS SECURED IN TEACHING MATHEMATICS 5A	39
VII. RESULTS SECURED IN TEACHING MATHEMATICS 5B	47
VIII. SUMMARY OF RESULTS SECURED BY INDIVIDUAL TEACHERS	55
IX. INTERPRETATION OF THE FINDINGS	61
X. LIMITATIONS AND APPLICATION OF THE METHOD USED	67

II

THE VALIDITY OF CERTAIN INSTRUMENTS EMPLOYED IN
THE MEASUREMENT OF TEACHING ABILITY

	PAGE
I. THE STATEMENT OF THE PROBLEM	73
Methods of evaluating teaching. Need for objective measurement. Purpose of the present study.	
II. THE TEACHING POPULATION, THE PROCEDURE, AND INSTRUMENTS OF MEASUREMENT	79
The population studied. Measures of pupils. Measures of teachers.	
III. THE CRITERIA OF TEACHING SUCCESS EMPLOYED IN THIS INVESTIGATION	85
An ideal criterion. The criteria employed. Justification of the criteria.	
IV. THE VALIDITY OF CERTAIN MEASURES OF TEACHING ABILITY WHEN GAIN IN PUPIL ACHIEVEMENT IS USED AS CRITERION	90
The measures studied. Equivalence of gains at different levels. Differences between cities. Curricular validity of the Stanford Achievement Test. Factors affecting the coefficients of correlation. Correlation of teacher measures with gain. Relation of intelligence to gain. Summary.	
V. THE VALIDITY OF SEVEN RATING SCALES WHEN GAIN IN PUPIL ACHIEVEMENT IS USED AS CRITERION	109
The measures studied. Correlation of rating scales with gain in achievement. Influence of intelligence.	
VI. THE VALIDITY OF THE NINETEEN INSTRUMENTS OF MEASUREMENT WHEN FIVE COMPOSITE CRITERIA ARE EMPLOYED AS MEASURES OF TEACHING SUCCESS	118
The composite criteria. Procedure. Gain in raw score on Stanford Achievement Test. Gain in accomplishment quotient. General merit ranking. Seven teacher rating scales. Six tests of teacher qualities. Three additional measures of teaching ability. Relative merits of the several measures. Summary.	

TABLE OF CONTENTS

xix

	PAGE
VII. THE VALIDITY OF THE SEVERAL MEASURES OF TEACHING ABILITY WHEN A WEIGHTED COMPOSITE IS USED AS CRITERION	130
The new criterion. The problem. Validity of the measures. The predictive value of each of the variables as compared with general merit ratings. The predictive value of a selected team of tests. Conclusions.	
VIII. GENERAL SUMMARY	138
Conclusions. Further generalizations.	

III

PUPIL ACHIEVEMENT AND THE NS TRAIT IN TEACHERS

ACKNOWLEDGMENTS	144
I. PURPOSE OF THE STUDY	145
A. EVIDENCE FROM PREVIOUS STUDIES	146
Pupil achievement and the training and experience of the teacher. Pupil achievement under student teachers. Pupil achievement and the ability of the teacher. Pupil achievement and the professional information of teachers.	
B. CRITICISM OF TECHNIQUES USED IN PREVIOUS STUDIES	150
The meaning of "teaching ability." The meaning of "pupil achievement."	
II. FINDING A MEASURE OF PUPIL ACHIEVEMENT	156
Pupil gains. The reliability of pupil gains. Pupil achievement in relation to ability.	
III. SELECTION OF ITEMS FOR THE TEACHER TEST	160
Paired comparisons. Dislike for teaching. Ease of learning difficult tasks. Recognition of the importance of tasks. Experience versus pre-service training. Recognition of important forms of behavior. Indecision response. Professional pride. Race prejudice. Personal adaptability. Unmitigation in assertions. Disregard for superstition. Distractability.	

	PAGE
IV. REVISING THE SCORING SCHEME	170
The teachers studied. The criterion groups. Weighting the scores. The index of importance. Index of independence. The direction of scoring. The final weights. Calculations and checks. The reliability of the sum of weighted NS scores.	
V. VALIDATION OF THE INSTRUMENT FOR MEASURING THE NS TRAIT	188
Selection of the subjects. Measures of validity.	
VI. PUPIL ACHIEVEMENT AND THE NS TRAIT IN TEACHERS	190
Source and collection of data. Reliability of measurement and of sampling. Statistical technique. The results found. Interpretation of results.	
VII. SIGNIFICANCE OF THE FINDINGS	197
The meaning of the NS trait. The meaning of the pupil achievement score.	
VIII. SUMMARY	202
APPENDIX: The Information Booklet	205