Columbia University

Contributions to Education

Teachers College Series

Qualities Related to Success in Teaching

By
FREDERIC BUTTERFIELD KNIGHT, PH.D.

Teachers College, Columbia University
Contributions to Education, No. 120

Published by
Teachers College, Columbia University
NEW YORK CITY
1922

ACKNOWLEDGMENTS

THE writer congratulates himself upon the fact that for many happy months he was a student of Dr. E. L. Thorndike, of Teachers College, to whose genius and leadership an ever-growing army of students owe their greatest inspiration. He wishes to thank the teachers whose unusual patience and forebearance made the gathering of the data for this study possible, and to express also his deep obligation to Dr. George D. Strayer and Dr. William C. Bagley, of Teachers College.

PREFACE

The only variation from current practice in the organization of the subject matter of this study is the changing of the position of the conclusions. The writer has stated his conclusions at the very beginning of the study. This was done to enable the busy reader to see at a glance the general trend of the study.

As the practical school administrator knows, the gathering of the kind of data upon which this study is based is a matter of some difficulty. It is the kind of data, however, which we must have if science is to aid in the selection of teachers.

Some psychologists may contend that an analysis of teaching rather than the correlation of observable facts with varying amounts of success of actual teachers is the only correct method of determining what tests will distinguish good from poor teachers. No one would deny the value of sagacious insight into any problem of human engineering. So far neither the analysis method nor the correlation method has done very well in practice on the practical job. This study is based on the correlation method. Its shortcomings should not be confused either with the logical soundness or with the practical superiority of test construction on a basis of correlation between test scores and performance.

FREDERIC B. KNIGHT.

CONCLUSIONS BASED ON THIS STUDY

This thesis deals with the problems of isolating the significant and measurable qualities of effective teaching and the methods of measuring these qualities. It is a continuation of similar studies of which the work of Meriam was the first. A rating of 153 high-school and elementary-school teachers was obtained by having the teachers rate each other for the quality of general teaching ability and other traits. While it may be said that the teachers knew each other only in a social way and, therefore, could not rate each other for general teaching ability, the data show that an adequate rating can be procured by this method.

The statistical treatment of the data shows that:

1. Chance halves of the mutual ratings of the teachers correlate with each other $+.899, \pm .01$.

2. The mutual ratings of the teachers correlate with the ratings of the supervisors $+.962, \pm .001$.

3. The mutual ratings of the teachers correlate with pupils' estimates $+.681, \pm .05$.

4. There is also substantial agreement among the mutual ratings of teachers, when they rate for specific traits, such as intellectual strength and skill in discipline. The average correlations between chance halves of the ratings are respectively $+.879, \pm .016$, and $+.838, \pm .023$. These correlations are evidences of the ability of teachers to rate each other.

The ratings for general teaching ability, secured in this way, were used as measures of teaching merit, against which objective facts were correlated. The correlations between general teaching ability and age, amount of experience, quality of handwriting, intelligence as measured by test, major academic interests, normal-school scholarship, amount of professional study during active service, and ability to pass a professional test have been secured. The correlations are too low to warrant one in using these factors for prognostic purposes, except ability to pass a professional test ($+.541$), normal-school scholarship ($+.153$) and intelligence ($+.108$). By using the coefficient of partial correlation we find, in

the case of elementary school teachers, that, the factors of intelligence and normal-school scholarship being constant, there is a mutual relationship of +.57 between ability to teach and ability to pass a professional test. Professional tests may be used to estimate probable success in teaching. The amount of professional study accomplished during active service is also indicative of success in teaching. The number of teachers who had accomplished professional study of this sort was too small in the groups which were studied to allow an accurate determination of the degree of significance that professional study has.

In the case of high-school teachers, intellectual differences, as determined by mental tests, appear to be significant. For the selection of high-school teachers the use of mental tests would be of value.

These data, as a whole, may be interpreted to mean that the general factor of interest in one's work becomes the dominant factor in determining one's success in teaching. The reasoning which leads to this conclusion is not straightaway, for we have not as yet objective tests of interest. We do know, however, that other measurable traits, either alone or in combinations, are not adequate explanations of teaching success. With our present knowledge it is reasonable to suppose that genuine interest in one's work accounts for a large part of teaching success.

In the second part of the study data are presented which show the spread of general estimate to particular traits, when judgments or ratings are made. For example, when a judge attempts to rate a teacher in some particular trait, his rating is a defense of his general estimate of that teacher, as well as a rating of the trait under consideration.

The mutual judgments of teachers for the trait, intellectual ability, correlate with their judgments of general teaching ability +.935, ± .014.

The mutual judgments of teachers for the trait, skill in discipline, correlate with their judgments for general teaching ability +.789, ± .001.

The mutual judgments of teachers for the trait, skill in discipline, correlate with their judgments for intellectual ability +.863, ± .080.

It would be difficult to hold that these correlations represent the true relationships which exist between these pairs of traits. The

presence of a large factor of spread of general estimate accounts best for the size of these correlations.

A study of the correlations between the ratings of 126 teachers in a New York school system for 15 traits showed that 105 of the 120 correlations studied could be accounted for by chance variation from an average correlation, even if a perfect, or a 100 per cent, spread of general estimate was present.

A study of the correlation between qualities of teaching as presented by Boyce in his work, published in the *Fourteenth Yearbook of the National Society for the Study of Education*, shows that 85 per cent of the correlations come within a range of $\pm .150$. These facts can be satisfactorily explained only when a factor of spread of general estimate is allowed.

It seems fair to conclude, therefore, that in judging particular traits general estimate influences the particular estimate to such a degree that judgments of particular traits are in themselves of little practical use.

CONTENTS

CHAPTER I

INTRODUCTION

THIS thesis[1] lies in the field of research which is concerned with methods of rating teaching, of determining the significant factors in teaching ability, and of measuring objectively such factors.

This field of research is by no means a virgin one nor is it one of academic interest only. Practical school administrators have no more important and, at present, no more troublesome problems than those which are grouped around the technique of selecting and rating teachers.

Actual isolation of such factors as intellect and temperament, which are indispensable to successful teaching, and the discovery of a method determining whether a prospective teacher possesses the indispensable qualities of a good teacher would be a boon to school administrators.

During the past fifteen years educators and psychologists have given their earnest attention to this series of problems, on which a great deal has been written and on which much research work has been done. Three studies have been selected to show the general development of attempts which have been made to find solutions to different phases of the personal-management problems of our public schools.

MERIAM'S STUDY

Dr. L. L. Meriam, in a research study, *Normal School Education and Efficiency in Teaching,* published in 1906, Teachers College Contributions to Education, No. 1, Chapter IV, presented data which were used to discover the correlation between teaching efficiency and scholarship in the normal school.

"This is the problem," said Dr. Meriam. "Is the efficient teacher the proficient scholar? To what extent is he so in each of the subjects of the normal-school course? In other words, does the one who stands high among fellow-teachers stand relatively high among fellow-students in the work preparatory to his

[1] All data used in this study are on file at Teachers College, Columbia University, New York City.

teaching? Such a study of mental relationships is in itself a study of causes. If it be found a rule that efficiency in teaching follows proficiency in scholarship, then, other things being equal, the latter may be considered a vital contribution to the latter. And this is our present purpose: to discover, so far as possible, what elements enter into the making of a capable teacher. Corollary questions are: To what extent does proficiency in scholarship mean efficiency in teaching? . . ."

In Dr. Meriam's research study an admirable attempt was made to find out the relative teaching ability of a large number (1,185) of normal-school graduates. Equally careful work was done to determine the relative normal-school success of these graduates. Meriam had no accurate measure of teaching efficiency and no reliable measure to equate the amount of success in one school system with the amount of success in another. He encountered the same difficulty in interpreting normal-school marks as measures of scholastic accomplishment. Great statistical ingenuity was shown by Dr. Meriam and his results, by all odds, were the most dependable at the time of the publication of his thesis.

The correlation between normal-school standing and ability or success in the field was found to be so surprisingly low that differences in scholarship among students in the normal schools seemed to bear a negligible relation to future differences in teaching ability. Meriam found that practice teaching during normal-school training was slightly prophetic of the quality of teaching which should be expected after graduation. Examinations containing professional subject-matter did not appear to furnish a significant index of an individual's ability to teach.

The statistical difficulties of Meriam's work should not blind one to its value. It clearly stated the problem of correlating teaching ability with factors which are more or less objective and measurable. It developed a technique of research that was sound in theory. It exercised much influence in taking the problem of teaching efficiency from the field of opinion and discussion and in placing it, where it properly belongs; namely, in the field of research and objective measurement.

Meriam's more important findings are expressed as coefficients of correlation between teaching efficiency and scholarship in normal-school studies. These he reports as follows:

Correlation between Teaching Ability and Practice Teaching	+.39
Correlation between Teaching Ability and Psychology	+.37
Correlation between Teaching Ability and History and Principles of Education	+.28
Correlation between Teaching Ability and Method Courses	+.29
Correlation between Teaching Ability and Academic Courses	+.22

Meriam's data also support his conclusion that, after the first year of teaching, experience, as such, has little if any influence on the improvement of teaching efficiency.

ELLIOTT'S STUDY

In 1910 another treatment of this general subject was published which deserves notice. Dr. Edward C. Elliott presented to the second annual convention of city superintendents in Wisconsin "A Tentative Scheme for the Measurement of Teaching Efficiency." This score card has been revised in detail, but the first scheme included all the essential factors. Elliott stated these three propositions which were of more than temporary importance:

"Is it possible to devise and to apply to the teaching process, impersonal, quantitative standards, whereby the relative worth and efficiency of teachers may be determined more justly and with greater precision than under the ordinary practices of the day?

"Does not the effective organization, administration, and supervision of public schools require that the conditions and results of the teachers' work be subjected to measurements of a quantitative rather than a qualitative nature?

"Is it possible for the present generation to make any reliable and satisfactory conclusions concerning the direction and rate of educational progress without standards of value resting upon a quantitative basis?"

The scheme divides teaching efficiency into seven sections and to each section assigns a weight or value. The scheme, in summary, is here reproduced:

I. Physical Efficiency	12	points
II. Moral-nature Efficiency	14	"
III. Administrative Efficiency	10	"
IV. Dynamic Efficiency	24	"
V. Projected Efficiency	6	"
VI. Achieved Efficiency	24	"
VII. Social Efficiency	10	"
Total	100	"

The value of this scheme is that attention is directed to particular traits and that diagnosis of teaching merit is stimulated. The suggested values are, of course, matters of opinion. The assumption that analysis of the teacher and of the judgment of particular qualities, studied in isolation, can be made is highly questionable.

BOYCE'S STUDY

In addition to Meriam's study the only other research of extensive nature is that made by A. C. Boyce[1] and published under the title "Methods of Measuring Teachers' Efficiency," Part II of the *Fourteenth Year-Book of the National Society for Study of Education.* Boyce obtained the rating of a great many teachers for general merit and for specific qualities. Then, by a method of correlation, he worked out the relative significance of the qualities. There are many technical improvements in this study over that of Meriam's, but the general procedure is the same.

For fifteen years the teaching profession has been sensitive to problems of recruiting new members. As yet, however, no one knows the exact formula for success in teaching. The complexity of personality and character and the many-sidedness of teaching have continually baffled useful analysis. We know that several measurable traits are *not* essential to successful teaching, but we do not know what traits *must* be present in superior instructors. The inspiring advance in the application of psychological methods to the selection of clerks, stenographers, machine operators, and fliers in industry together with similar success in vocational guidance in professional education, such as engineering and dentistry, increases our confidence in the hope that before long psychology will enable school administrators to select teachers with frequency and size of error far smaller than prevails at present.

[1] For a discussion of this study, see the last part of Chapter V.

CHAPTER II

METHOD AND DATA INVOLVED IN THIS STUDY

An accurate rating of a sufficiently large number of teachers for general teaching ability must be obtained before any analysis of the significant qualities of teaching is possible. We must know who the *good* teachers are, who are the *poor* teachers and who are the *fair* teachers, before it is worth while to attempt to find out what facts are pertinent in judging their teaching skill. After we get a group of teachers who we know differ among themselves in general teaching ability, by certain amounts or units, then we may proceed, by a method of correlation, to find out what facts about them are of prognostic or diagnostic value.

Such a rating of general teaching ability for 156 grade and high-school teachers who were at work in the public schools of Towns A, B, and C, in Massachusetts, during the school year 1918–1919, has been obtained. There were six groups of teachers. Three of these groups were the grade teachers in Towns A, B, and C, and three were the high-school faculties in these towns. The number of teachers in each group follows:

Town A	Grade teachers	53
	High-school teachers	15
Town B	Grade teachers	35
	High-school teachers	13
Town C	Grade teachers	30
	High-school teachers	10

Three separate ratings for general ability in teaching were obtained for each group. One rating was secured from the supervisors in each system for their respective teachers. Another was secured by the mutual judgments of the teachers themselves of each group. Another was secured by a consensus of pupils' opinions.

In general the method which was used in deriving the ratings was to have the several judges rate each teacher in the group relative to the other members of the group for the broad quality *general ability as a teacher*. The theory which underlies this method is this: Where direct measurement in terms of amount is impossible, measurements by relative position in a series may be so controlled that possibly as exact and as true ratings may be ob-

tained as if units of amount had been used. It is assumed that the amount of difference between two teachers who have been thus judged will depend on the ease with which the differences are observed by competent judges.

In using this general method of rating teachers for teaching ability, we have taken it for granted that the good teacher is the one whom competent judges rate as *good*. We hold throughout this study that the poor teacher is the one whom the judges have rated as *poor*. These hypotheses will presumably be acceptable to those who are familiar with the theory and practice of social measurements. It may be admitted that one could question the final truth that the opinions of any number of judges, however competent and harmonious they might be, necessarily establish the facts of teaching merit. Thus the really good teacher, it might be held, is the one who gets her pupils on fastest.

To determine how much the progress of pupils is due to any one teacher is not possible by any method or information that is as yet available. Even to measure a pupil's total progress, much less the total progress of a class, is as yet a little venturesome.

It might also be held that the amount of development of character and morality in the pupils is the only test of good teaching and that what others think about the teacher is really irrelevant.

To hold, on the other hand, that competent judgments of teachers, when properly combined, will give a very useful and approximately true rating, as well as probably the best rating method that is now available, is only common sense. This rating method is entirely defensible.

The good lawyer, after all, is the one who is considered a good lawyer by fellow-members of the bar. The poor dentist is the one to whom no other dentist would go or recommend anybody else. The great preacher is the one who attracts visitors. The good teacher is the teacher who is thought to be good.

Where differences in skill among employed people must be determined, judgments in terms of *better than the average, poorer than one's associates*, and similar expressions, are useful measures of ability. Of course, the final validity of the judgments may be lessened by the presence of constant error in the opinion that it offered, or by the incompetence or paucity of the opinions expressed, or by the failure properly to combine the judgments after they are obtained.

Of the three ratings,—by the teachers themselves, which is labeled "A," by the supervisors, "B," by the pupils "C,"—we shall take up first the ratings of the teachers which are indicated by the judgments of their fellow-teachers.

PROCESS OF RATING TEACHERS, BASED UPON THE TEACHERS' ESTIMATES

Step 1. Teachers' meetings for each group were called and the teachers were asked to rate each other for general teaching ability, using the relative-position method. The ratings were not in terms of *good, fair, poor,* because what one teacher might consider *good,* another teacher who was more critical might consider only *fair.* This type of difference might run through the series of judgments.

The ratings were not secured in terms of *how much below the best teacher you have ever known,* or the equivalent expressions, for errors of an obvious nature are bound to creep into any such rating system. The ratings were all given in terms of relative position within the group itself. Thus, when the grade teachers of Town A rated each other, every teacher placed in order of merit all the teachers in the Town A group. The amount of difference between the teachers in the final rating was determined by combining all the judgments of the teachers. Each member of the six groups of teachers, while in a teachers' meeting, rated those in the group to which she belonged in a similar fashion and under similar conditions with the same instructions. The instructions which were given to the teachers follow:

INSTRUCTIONS TO TEACHERS

On this sheet you are requested to give certain ratings of each teacher in the list, including yourself. Please rate every teacher and please be absolutely frank in your ratings. You need *not* sign your name. Nobody will ever know how you or anybody else rated him. No personal use will ever be made of any of these ratings. They will be used in a purely scientific study to determine the significance of age, education, early interests, etc., etc., for success as a teacher. The names will all be cut off and destroyed as soon as the different items in the inquiry have been numbered to fit the ones to whom they refer. Also, do not feel disturbed because in each respect somebody has to be rated lowest. These ratings are all relative, and the lowest teacher in the group may well be of very great ability. Please be sure to record ratings, even if they seem to you to be little better than mere guesses. The opinions of twenty men give a useful rating, even if any one of the twenty taken alone is almost worthless.

On the sheet is a list of the teachers. Choose the teacher of greatest *teaching ability* and write 1 after his or her name in Column 1. Choose the teacher next below in teaching ability and write 2 after his or her name in Column 1. Write 3 after the name of the one next below in teaching ability, and do so for 4, 5, 6, etc. If two or more seem absolutely equal in teaching ability give them the same rating.[1]

After the teachers had read the instructions carefully a few minutes were allowed them for asking any questions that might occur. When it was clear that the teachers understood what was wanted of them, they proceeded with the rating. No names were signed to the rating sheets. It was evident that honest and sincere opinions were expressed. The resultant ratings of each of the six groups of teachers were then examined. Those sheets which were incomplete or did not sufficiently distribute the ratings were discarded. This lack of usable material was not at all great.

The teachers found that rating each other was a method of polite gossip and was evidently more or less enjoyable. For each set of teachers sufficient material was obtained. The spread or range between the poorest and the best teacher was large. In many cases it was as great as the number of teachers involved. The number of useful ratings (97) were distributed as follows:

Town A	Grade teachers	30	ratings
	High-school teachers	14	"
Town B	Grade teachers	16	"
	High-school teachers	10	"
Town C	Grade teachers	18	"
	High-school teachers	9	"
	Total	97	"

Step 2. Each set of ratings was then divided into two halves by chance drawings. Each half has been treated separately throughout this study. These halves will be referred to as Group A and Group B. The carrying of two groups makes corrections for attenuation in the correlations and shows also the reliability of the judgments.

A transcript (see Table I) of fifteen of the ratings for general teaching ability of Town A grade teachers has been made. These ratings compose one group (Group B) of the mutual judgments which is treated later to get a single rating of teachers. The

[1] The complete instructions are given on pp. 46-48.