



A Proposed Index of Efficiency in Teaching United States History

B. R. Buckingham

To cite this article: B. R. Buckingham (1920) A Proposed Index of Efficiency in Teaching United States History, *The Journal of Educational Research*, 1:3, 161-171, DOI: [10.1080/00220671.1920.10879038](https://doi.org/10.1080/00220671.1920.10879038)

To link to this article: <http://dx.doi.org/10.1080/00220671.1920.10879038>



Published online: 15 Dec 2014.



Submit your article to this journal [↗](#)



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

Journal of Educational Research

Editorial Office, Urbana, Illinois

Editor-in-chief B. R. Buckingham, University of Illinois

VOLUME I

MARCH, 1920

NUMBER 3

A PROPOSED INDEX OF EFFICIENCY IN TEACHING UNITED STATES HISTORY

B. R. BUCKINGHAM

University of Illinois

When a supervisor enters a classroom to inspect the work and rate of the teacher, he carries with him a set of more or less variable personal standards. He may also fortify himself with a schedule of qualities on the basis of which he will perhaps believe that he is scientifically evaluating the work of the teacher. We have a number of score cards which set up such captions as personality, voice, neatness, method, control, and the like. These are by no means satisfactory. Teachers possessing these qualities to a high degree may nevertheless be poor teachers. Personality as a producing agency, neatness functioning, voice used to get results, method which does the business, and control which facilitates work—these qualities, in other words, not as possessions of the teacher but as productive of changes in pupils—are touchstones of good teaching.

We must, therefore, seek for teaching efficiency in the changes produced in children rather than in the qualities possessed by teachers. We are not, however, to suppose that we have greatly simplified our problem when we have thus defined it. The results of teaching are numerous and far reaching. We can never hope to measure all of them, for they multiply to the end of life.

Yet we are not without means of measuring with a considerable degree of accuracy the efficiency of teaching. What we need to do is to find some one major result the simplicity of which makes its apprehension relatively easy and the relationship of which to

the totality of result is close. If there were a perfectly definable and generally recognized relation between the length of the left forefinger and the stature of a human being, and if it were difficult to measure stature but easy to measure finger length, then the height of human beings might very probably be expressed not directly but indirectly through the measurement of the left forefinger. Countless uses are made of the principle of directly measuring a magnitude in order that another magnitude related to it may be indirectly measured. A printer often counts the material in a pile by measuring its height or weight. Taxes have been assessed according to the number of windows in the houses because the value of the house and hence the wealth of its owner were thought to be roughly proportional to the number of windows. We weigh the air by measuring the height of a column of mercury which it will sustain; and the weight of the air is itself functionally related to the weather conditions which we are thus, in turn, able to estimate. Again, we directly measure the expansion of a fluid in a tube and thereby indirectly measure temperature.

In the matter of teaching it is necessary in the first place to seek our index of efficiency among the results of the teaching. In the second place, it is desirable that this index shall be of a simple and easily measurable nature.

Some of the results of teaching are being measured objectively by means of standardized tests and scales. Opposition to educational measurement as such is all but silenced. Not many voices are now being raised against it on principle. Numerous critics, however, assail its practical applications and they do so at many points and in many ways.

It is not unusual, for example, for the critic to contend that certain subjects do not lend themselves to measurement. This assumption is a dangerous one, and when the critic is so rash as to make his objections apply to a particular subject he is likely to be refuted by the next educational periodical or bulletin. In 1913, when measurement had been applied for two or three years to handwriting and arithmetic, it was pretty generally believed that these two subjects might be susceptible to such treatment, but that one could go no further. In particular, it was roundly asserted by a speaker, and cordially approved by his hearers, that spelling was a subject that would never yield to measure-

ment. Yet everyone knows that to measure the spelling ability of pupils is one of the first steps of the novice in measurement. Similarly, nobody thought that ability in reading could be measured, yet the sale of millions of copies of reading tests now assures us that its measurement is at least being widely attempted, and that people believe they are securing reasonably reliable results.

In the case of United States history, there is evidently a phase of the subject which lends itself admirably to measurement. The method of expressing the difficulty of the elements in any test has been carefully worked out. It has, therefore, become perfectly feasible to apply this method to any subject, the difficulty of whose elements can be determined. One definition of difficulty—not the only one, to be sure, but one which is especially significant—is the proportion of correct responses to the task in question. When did Columbus discover America? Who was president of the Southern Confederacy? What two probable European ventures was the Monroe Doctrine set forth to forestall? These are matters whose difficulty on the basis of the proportion of correct responses may be as accurately determined as the difficulty of the number combinations or of the spelling of words.

"Arrange the following events in the order in which they occurred" says one author of a history test. Since there is but one order of occurrence and that order is a matter of record about which there can be no dispute, this question is likewise one whose difficulty can easily be determined. Adopting one of the devices of the Army Test whereby a number of ends of sentences are given with instructions to the subject to underscore the right one, another author gives questions of this type: "The Huguenots first settled in *Baltimore, Boston, Charleston, New York, Philadelphia*" Or again, "The Patroons were Dutch *fishermen, fur-traders, land owners, miners, preachers, teachers.*" Again there is only one correct response and its correctness is not questioned by qualified persons. Many other types of questions might be listed—questions the responses to which are either undoubtedly right or undoubtedly wrong. Asking the names of men to be connected with events; requiring the selection, from a given list, of the policies or events favored by a given party; the association of events with dates, of causes with effects, of men with policies,

with principles, or with each other; the classification of men or events in various accepted ways—these are some of the types of exercises now being used. As to the form which answers may take, there is also an increasing richness and variety. One of the earlier devices was to print a question in a box and require the answer to be written in it. Numbers are used to indicate facts, particularly facts regarding the sequence or importance of events. Instead of requiring the pupil to write out the names of men, events, measures, policies, etc., a number of these are printed and check marks are required. The true-false type of test has been introduced.¹ It consists of a series of statements, some of which are true and some of which are false and the pupil is required to indicate the truth or falsity of each statement. The completion type of test has been with us a long time and has proved exceedingly useful.

But in order to be certain that responses to questions are either right or wrong, the questions must, clearly, be questions of fact. One may admit the ease with which such questions may be rated and the cleverness of authors in arranging fact tests and yet remain in doubt as to the value of the results of such tests as measures of ability in history or of the efficiency of the teaching of history. The problem then resolves itself into this: granted that fact questions are easy to devise and easy to rate, what meaning can we give to the results? The critics of measurement have all along been saying that the measurers were formalizing education, that only the least valuable subjects or parts of subjects could be measured, in short, that the only appeal was to the memory or to a few simple skills such as those involved in handwriting.

There are those who seem to think that memory was given us as a human attribute only to be despised. In the vocabulary of the modern pedagogical writer, the word memory is only used in a deprecatory sense. In my own judgment, the case against memory has been vastly overstated. This opinion, however, will not be urged in this connection. But it will be apparent as we proceed that, even when we think we are appealing to a supposedly higher process, we may really be dealing only with a somewhat higher form of memory.

¹ McCall, W. A. "A new kind of school examination," *Journal of Educational Research*, 1:33-46, January, 1920.

Fact questions are those questions whose correct answers are facts. Observe, that the determination of whether a question is a fact question or not depends upon the character of the answer which is accepted as satisfactory. We may ask a child what he *thinks* were the results of the War of 1812; but if we only credit him with a satisfactory answer when he tells the actual results of the war, we are dealing with a fact question, no matter how we delude him or ourselves with the notion that he is fancy-free, because we have asked him what he *thinks*. A fact question appeals to the memory. It may be a date or a name, an incident or an event, a policy or a movement, a cause or an effect, an evaluation or a relationship. It may be bald and meaningless or it may be significant, and even surcharged with emotion.

It appears, therefore, that these fact questions—however rich they may be in meaning—are nevertheless capable of being objectively rated and easily scored.

If, therefore, it can be established that test scores on fact questions are directly related to the thing we call historical ability, we shall be justified in measuring historical ability by means of the ability to answer fact questions. This relationship may be mathematically established if we can secure from the same group of pupils two sets of measures: (1) scores on fact questions and (2) scores in historical ability. I am not sure that I know what historical ability is. Doubtless, psychologically speaking, there is no such thing as historical ability but only many different historical abilities; in other words, the ability is probably plural rather than singular. I presume, however, that we shall come somewhere near the conception of historical ability if we secure measures of ability to think correctly and to judge characters correctly in the field of history. At least in the most elaborate and carefully analyzed history test, namely that of Van Wageningen,² the assumption appears to be made that historical ability is rather fully covered when we have taken into account information, thought, and character judgment.

² Van Wageningen, M. J. *Historical information and judgment in pupils of elementary schools*. New York: Teachers College, Columbia University, 1919.

Van Wageningen, M. J. *American History Scales* (Information Scale, Series A and B; Thought Scale, Series A and B; and Character Judgment Scale, Series A, B, and L). Published by Teachers College, Columbia, University, New York, 1919.

Granted that this is true, two problems challenge our attention: First, the extent to which ability to think and to judge character in the field of history depends upon a knowledge of facts; and second, the extent to which thought and character-judgment questions (using those of Van Wagenen as typical) are really fact questions.

In 1915 I submitted certain fact questions and certain thought questions to some elementary school children in New York. A little more than a year later I gave a different series of fact and thought questions to elementary and high-school pupils at Madison, Wisconsin. I found a correspondence between the ability to answer these fact and thought questions which may be mathematically expressed by a correlation coefficient of about $+0.4$.³ I also found that if one child scores a unit higher than another on the fact questions, he would also, on the average, score 0.89 of a unit higher on the thought questions. From these facts, I derived an equation which would permit the probable score in the thought questions to be forecast from a known score in the fact questions with the chances even that the prediction would not vary more than two out of a possible twenty points from the score which would be obtained by actually administering the thought questions.

The far more extensive testing of Van Wagenen yielded even higher correlations. Between his fact scores and thought scores he found a correlation of $+0.80$ and between his fact scores and character judgment scores the correlation was $+0.78$.

The correlations reported by Van Wagenen are for seventh and eighth grades. I gave his tests to pupils in two high schools, one at Champaign, Illinois, and the other at Danville, Illinois, and obtained correlations between fact and thought scores of $+0.53$ and $+0.60$ respectively. Between fact and character

³ The reader who is unfamiliar with the correlation coefficient may get a rough and ready idea of its meaning by considering that if two series of measures—e.g., scores in fact questions and scores in thought questions—are perfectly and directly related, the correlation coefficient is $+1.00$. Under these circumstances the larger one measure is the larger the other will be. If the measures are not related to each other at all, the correlation coefficient is zero. Coefficients between 0 and $+1.00$ indicate closer direct relationship the more nearly they approach $+1.00$. A coefficient of $+0.4$ indicates a well-marked tendency for the two measures in question to vary together; but it also indicates a rather large number of exceptions to the tendency.

judgment scores the correlations were $+0.45$ and $+0.43$ respectively.

As between fact and thought scores, the relation, therefore, seems to be very close. My high-school results are not as high as Van Wagenen's seventh and eighth grade results because the thought test (a test which Van Wagenen used in grades as low as the fourth) proved to be too easy to distinguish thought ability among the more capable children. But if in grades for which these fact and thought questions are appropriate the correlation is $+0.80$, we are justified in saying that the two tests tell much the same story—in other words, that, after having given the Van Wagenen information test, we shall gain little by giving his thought test.

Nothing, therefore, either in Van Wagenen's testing or mine justifies one in attempting to measure all the desirable outcomes of history teaching. Fact scores seem to be surprisingly sufficient. Moreover, thought and character judgment are hard to rate; and one rater cannot be altogether consistent either with another or with himself at another time. Accordingly, any added contribution which the testing of such abilities may make to our estimate of "general ability" in history will not only be small but will be measured with uncertainty.

To be sure, thought and character judgment do not represent all the desirable outcomes of history teaching. We wish our pupils to grasp the meaning and significance of historical situations, to apprehend changing standards of life, to judge the reliability of historical material or evidence, to distinguish between statements of fact and statements of opinion. This list of aims is taken almost verbatim from Van Wagenen; and doubtless it could be extended. But as to these, may they not all be shown to rest upon a knowledge of facts? How may one "grasp the meaning and significance of historical situations" unless one has at hand the facts about the situations and the additional facts which will give them "meaning and significance"? How may one "apprehend the changing standards of life" unless one knows what the standards were before and after the changes took place? How may one "judge the reliability of historical material or evidence" unless one knows the material and the evidence in question and the facts or criteria on which their

reliability depends? Finally, how may one "distinguish between statements of fact and statements of opinion"—except so far as the form of the statement itself furnishes the distinction—without knowing the facts in the case?

These queries lead to the statement that a knowledge of facts—information, as Van Wagenen calls it—is symptomatic of higher abilities in history. The evidence supporting this view arises not only from the fact, as shown above, that pupils who know facts will also think and judge well but also from two further conditions. First, we may extend the term "facts" to include exact information of any nature in the field of history—information not only as to dates, persons, and places but also as to causation, relationship, political and social movements, economic developments, religious beliefs, wars and rumors of wars, peace and the arts of peace. Second, knowledge of facts conditions higher abilities because many items usually included under these higher abilities are in reality essentially factual. Since Van Wagenen is the first test maker to present a thoroughly worked out scheme of items other than those of information, we may profitably consider some of them. I shall examine the first four questions of his Thought Scale, Series A, together with his acceptable answers to those questions.

The first reads as follows: "Before the steamboats were made people used to travel on the ocean in sail boats. Steamboats were not made until a long, long time after the European people came to make their homes in America. How do you think these early European settlers came to America?" The acceptable answer is "in sail boats" and it is a fact. Therefore the question is a fact question although introduced by the words, "How do you think."

The second thought question reads: "A little before the year 1500 the people of Europe were anxious to find a new way to get to India. Some people thought that India might be reached by sailing westward across the Atlantic Ocean. Columbus was one of these people. It was at this time that Columbus found America. What do you think Columbus was looking for when he found America?" Full credit is allowed for any of the following responses: "A short route to India"; "India"; "Western passage to India"; "Western route to India"; "Northwest passage

to India;" "New way to India;" "A way to India;" "An easy way to India." Here again the answers accepted are facts, and the question is, therefore, a simple fact question. The child is asked what he *thinks* Columbus was looking for, but evidently no thinking is expected.

The third question on the Thought Scale reads: "A hundred years ago it took a letter several days to go from New York to Boston. Today it takes only a few hours. Why do you think it took letters so much longer to go to Boston 100 years ago than it does today?" These answers are allowed full credit: "A hundred years ago there were no railroads;" "no trains or steamships;" "travel by stage coach or horses." Upon what, then, does the ability to answer the question correctly depend? Does it not clearly depend upon a knowledge of facts?

Question four: "The Northmen probably came to America as early as the year 1000, nearly 500 years before Columbus and the Cabots sailed from Europe. There is no record of any one else having come to America before the year 1000. By whom do you think America was first discovered?" Acceptable answers are: "The Northmen," "Vikings," and "Lief Ericcson." If we ask what enables a child to answer "Vikings" or "Lief Ericcson" are we not obliged to say that it is a knowledge of facts? The thought element in all these questions is only in the *form*—a form which employs the word "think."

Not all of the questions on the Van Wagenen Thought Test are so obviously factual, but most of them are; and some of those which are not, call for thinking which has no necessary reference to history.

I shall not examine in detail the questions in character judgment. Each consists first of a narrative and then of a requirement to underline three of ten adjectives which best describe some character or group of characters in the narrative. Reading and language ability are necessary and almost sufficient for success. Facts play no part except as they are given in the narratives; and no other products of instruction in history appear to be required.

Whatever, therefore, may be the desirable outcomes of the teaching of history, it does not appear to be necessary that a valid test should require reactions with reference to all of them.

It may merely require reactions which are symptomatic of them. In other words, we may distinguish between the content and method of teaching on the one hand, and the content and method of testing on the other. Many do not believe this; they regard tests as teaching devices, or at least as encouraging certain types of teaching. This is not always true, and may never be true in a content subject such as history. A test, to be a valid measuring instrument, must provide a reliable quantitative statement which varies in direct proportion to the ability or abilities which we seek to measure. It may be one of the complex of abilities. It may—theoretically at least—be something which is not a historical ability at all just as the expansion of a liquid in a column is not temperature. If it is supposed that by setting up fact tests we encourage formal fact teaching, this supposition may itself be in error. The fact test can be as meaningful as we desire to make it. As I have pointed out, our facts may far transcend those of dates, persons, and places. Again, it will be found that the mere teaching of formal facts will do little to enable a child to pass a well-constructed fact test in history because the right kind of facts are not learned by a dead lift of the memory but by discussion and reflection, by collateral reading and self-expression.

If, after giving a fact test, we should desire to know how capable the pupils probably would be in other ways, we may infer this with substantial accuracy *without giving any other test*. Suppose, for example, that an instructor has given the Van Wagenen information test and wishes to infer the composite ability of the pupils to think and to judge character. The teacher may do this provided the relationship has been established between scores in the information test and the combined scores in the thought and character judgment tests. The correlation coefficient expresses a relationship but not in usable form for this purpose. As a result of the testing carried on in the high schools at Champaign and Danville, we derived an equation by which the required inference may be made. This equation was $y = 72.8 + 0.9x$.⁴ A number of applications of this equation were made and most of them yielded results in close conformity with the actual test

⁴ Such an equation is called a *regression equation*. In this case, x stands for the fact scores and y for the combined scores in thought and character judgment.

results. The first pupil in our alphabetical arrangement had a score in the information test of 33. Substituting this for x in the equation we find that $y=103.5$. This then is the inferred score in thought and character judgment combined. The real score was exactly 100.

The prophesy, however, does not always work as well as this. The second pupil in the alphabetical arrangement scored 36 on the information test. When this value is substituted for x in the equation, it yields a theoretical score of 105.2. The real score was 114. Thus the prophesy is 8.8 units less than the actuality.

It will be interesting to have a general expression for the accuracy that a teacher may expect from the use of this equation. The equation only purports to give the score that on the average would be obtained by giving the thought and character judgment tests a large number of times. In any individual instance the chances are found to be even that the score obtained from the equation will differ from the score obtained by actual testing by not more than eight units. Since there are 144 units in the combined thought and character judgment tests, the chances are one to one (or we may say an even bet) that the score obtained from the equation will be within about 5 percent of the true result.

Thus we have given data and inferences to support the possibility of indirect measurement in history. The doctrine of indirect measurement requires us to find a magnitude which is easy to measure and which also varies closely with the more complex magnitude which we wish to measure but to which it is difficult to apply our instruments. It is maintained that fact questions—when the term fact is given its full and proper meaning—afford an easy means of measurement because of the unequivocal character of the responses and that these responses vary reasonably closely with what we are thus far able to set up as higher abilities in history. We have pointed out that the most notable attempt to measure two of these higher abilities has not succeeded in getting far away from a fact basis. It is also urged that a distinction be made between methods of teaching and methods of test making; and that if the type of testing which we adopt is thoroughly significant in the sense that it varies with the higher abilities, we shall be encouraging the training of these higher abilities when we use it.