Planck has unqualifiedly declared against it, and Einstein gave it up, I believe, some two years ago; and yet a quantum theory which fails completely to interpret or take any account of the most striking and the best established experimental fact which demands a modification of old theories, viz., the independence of the energy of emission of electron upon the intensity of the source, or, more generally, the inter-convertibility of β-rays and ether rays is, at best, a very impotent affair. If we are going to leave either of these two main groups of facts out of account I think almost any experimentalist would say that the first group (that having to do with the universal constant h) can most easily be spared; for if we could have radiant energy localized in space we might possibly account for all the experimental facts without having it emitted by a given source in exact multiples of something, but spreading ether pulses which contain energy in multiples of something are certainly wholly inadequate. They go but a short way toward accounting for the present experimental situation. In conclusion then we have at present no quantum theory which has thus far been shown to be self-consistent or consistent with even the most important of the facts at hand, and yet it looks as though one had to come, and when it comes I can scarcely believe that it will be one of the milder forms. That we shall ever return to a corpuscular theory of radiation I hold to be quite unthinkable. The facts of the static field alone seem to preclude such a possibility. But I see no a priori reason for denying the possibility of assigning such a structure to the ether as will permit of a localization of radiant energy in space, or of its emission in exact multiples of something, if necessary, without violating the laws of interference. That no one has as yet been able to do this can scarcely be taken as a demonstration that it can not be done. Fifty years ago we knew that such a thing as an atom existed, but we knew absolutely nothing about its structure, and it was customary to assume that it had none. To-day we know a great deal about the structure of the atom, but the position formerly occupied by it has been assumed by that thing which we call the ether. We know that there is a vehicle for the transmission of electromagnetic energy, but we know nothing whatever about its structure and it has been customary to assume that it has none. To deny the existence of this vehicle, which we have been in the habit of calling the ether, and to use the word "vacuum" to denote *all* the properties heretofore assigned to it by the experimentalist, viz., those of transmitting electromagnetic disturbances, is a bit of sophistry in which he is little interested. We seem to be on the eve of learning something more about the properties of this vehicle, call it by what name you will, than we have known heretofore. Certainly there has never been a time when physics offered such tasks to its followers as now, nor ever a time when it needed more and better brains applied to these tasks. It may be that "THOU art come to the Kingdom for such a time as this."

R. A. MILLIKAN

UNIVERSITY OF CHICAGO

EDUCATIONAL DIAGNOSIS[1]

UP till a score of years ago theories of intellectual and moral diagnosis suffered from two defects. They had not fully abandoned the notion that mysterious inner forces or agents existed—memory, attention, courage, imitativeness, constructive-

[1] Address of the vice-president and chairman of Section H—Education—American Association for the Advancement of Science, Cleveland, December, 1912.

ness, emulation, kindness, reasoning power and the like—each of which was a faculty or essence whose nature and degree of strength at any one time could be measured once for all. They had not even begun to abandon the notion that men were classifiable mentally into a rather small number of rather distinct types.

The early practise of diagnosis amongst psychologists, so far as there was any, consisted in inferring the condition of the individual's memory, imagination, power of attention, suggestibility and the like, from some facts in his behavior in which these magic powers "expressed" themselves, and in defining him as of this or that "type" on the basis of his divergence from the average in one or the other direction. Workers in schools who had progressed so far as to know of these practises either accepted their validity or distrusted them without knowing why they did so. The great majority of teachers continued to judge intellect, character and skill by the traditional means used and slowly improved by generation after generation since man could think and speak.

The active experimentation with objective "tests" of mental traits during the last score of years has shown that "goodness" of memory in the sense of a uniform power to hold all that is acquired, closeness of concentration in the sense of a uniform power to resist at will distractions of every variety and other similar general excellencies or defects, were myths. The measurements of correlations of the last decade have shown that "types" of attentiveness, of imagery and the like, and of intellect or character as a whole, either do not exist at all or are so complicated by intermediate conditions as to be of no service to thought or practise.

The experimentation with tests and the measurements of correlations began, however, with faith in the dogmas which they eventually disproved. The great majority of workers in the early days of tests assumed that certain formal general powers of mind existed, the strength of each being easily diagnosed by any one of its manifestations. Learning ten digits or nonsense-syllables was a test of "memory." He in whom the words violin, whistle, bell and brook aroused images of sound was thereby classified apart from him in whom they aroused images of sight. "Endurance" was measured by the ergograph, or by addition, or by the rate of tapping. Idiocy, imbecility and feeble-mindedness were three real entities, with symptoms awaiting our discovery. Even to-day much of this expectation that the intellectual and moral condition of an individual can be adequately described in adjectives, and manifests itself by the clear presence or absence of symptoms in the way that measles and smallpox do, remains.

Experimentation began to cure itself of these traditions as soon as it began to test the tests themselves. As fast as it abandoned the habits of assuming *a priori* what a certain fact in behavior signified concerning the individual's constitution, and sought instead to discover by exact—that is, quantitative—observation just what constitutional features it actually did go with, the older easy but false diagnosis was exposed and the basis for present and future achievement was laid.

Such studies of symptoms and tests are roughly of two sorts, those in which two distinct groups are compared in respect to the trait in question and in respect to the various possible symptoms of it, and those in which a long series of groups, each differing but little from the next in order, are so compared. In both cases certain principles of method are very desirable, almost necessary, because of the continuous grada-

tion of men in mental traits up and down from a common mediocre condition and the wide variability of the same person's performance in the same test.

In the case where two groups are compared in respect to some condition (call it T) and in respect to some possible symptoms thereof (call them S_1, S_2, $\cdots$, S_n) it is almost imperative that the amounts of T, S_1, S_2, etc., possessed by *each individual* be at least roughly determined. Unless this is done we can not tell how symptomatic of T each S is, nor compare the significance of these S's with others tested in some other investigations; we can only say of each of them that it is or is not, in some undefined degree, symptomatic of T; and perhaps that it is more or less so than others of the S's tested in the particular investigation. For example, in testing tests for general mental ability, Mr. Terman, who merely chose five very bright boys and five very dull boys without determining how bright or how dull they were, has left all his results undefined in quantity and incomparable with any one else's. An otherwise admirable study has thus had far less influence on educational diagnosis than it should have had. It is also desirable that the differences within either group in respect to T be small relatively to the difference between the central tendencies of the two groups, and that the extent to which each S separates the first from the second group be stated in an exact and commensurate quantity.

Unfortunately, the great majority of studies of sane *versus* insane, ordinary *versus* feeble-minded, bright *versus* dull, color-blind *versus* color-keen, musical *versus* non-musical, and the like, have left one or more of these three requirements unfulfilled.

I can not quote a study that is beyond reproach in all these respects, but the following facts from Dr. Simpson's test of tests of general intellectual ability will illustrate them roughly. Dr. Simpson defines the first of his two groups by the fact that they were all teachers or graduate students in a university. The group compared with them were all "men in New York City who had never held any position demanding a high grade of intelligence," who spoke "English as their mother tongue."

> Two of them were persons earning comfortable livings for their families, but men recognized by their associates as being dull. Eleven others were staying at the Salvation Army Industrial Home at the nominal salary of $1.00 per week in addition to board and room, until work could be secured. One of these held a somewhat responsible position at the time, acting as assistant superintendent of the home. He stood high in the most significant tests. The remaining seven were found in a mission on the Bowery where they were being helped somewhat until they could find employment.

As a measure of the extent to which each symptom or test separates the "Poor" from the "Good" groups, the degree of overlapping is taken. For example, the facts for four tests—giving the opposites of words, memorizing lists of words, marking the A's on a sheet of printed capitals, and discriminating lengths—were: The point reached by 50 per cent. of the "Good" group was reached by none of the "Poor" group in the opposite test or in memory of words; by 15 per cent. of them in the A test; and by 33 per cent. in the case of discrimination of length. The point reached by 76½ per cent. of the "Good" group was reached by none of the "Poor" group in the opposite test; by 5 per cent. in the case of memory of words; by 20 per cent. in the case of the A test, and by 62 per cent. in discrimination of length. The percentages of the "Poor" group reaching the points reached by 88 per cent. and by 94 per cent.

of the "Goods" are shown similarly in Table I.

TABLE I

Percentages of the "Poor" Group Reaching Certain Points of the "Good" Group's Ability

Percentages of the "Good" Group Reaching Certain Points on the Scales	Percentages of the "Poor" Group Reaching the Corresponding Points			
	Opposite Test	Memory of Words	A Test	Dis. of Length
50 per cent.	0	0	15	33[2]
76½ per cent.	0	5	20	62[2]
88 per cent.	0	10	25	73[2]
94 per cent.	0	10	25	88[2]

It is obvious that the opposite test is highly symptomatic of something which differentiates these two groups; that the tests in discrimination of length are not; and that the four tests are—in order of merit as symptoms—in the order named, giving opposites, memorizing words, marking A's and comparing lengths.

When a series of groups, each differing slightly from the next in order in the amount of the T or S taken as the basis of comparison, are used, the method of testing a symptom's significance is by its *correlations*—by the closeness of the correspondence of the rankings of the individuals in the S's with their rankings in the T. Thus twelve of the "Good" group in the preceding illustration were ranked, each by all[3] the rest (they being somewhat acquainted with the others' abilities), for general intellectual ability. It is then possible, by means which I will not rehearse here, to measure the closeness of correlation or correspondence between the measure for each man that would have been so obtained had this "imputed intelligence" been determined from 12,000, instead of 12, judges of each man, on the one hand, and the measure obtained for him by repeated testing with any of the tests. These correspondences as calculated for the four tests mentioned were: 96 hundredths of perfect correspondence in the case of the opposites test, 93 hundredths in the case of the memory of words, 21 hundredths in the case of the A test, and apparently zero in the case of discrimination of length.

This method allows a very delicate choice amongst tests, provided adequate data are at hand.

Suppose, for example, that arithmetical ability is defined as the ability required by a given selected set of a hundred arithmetical tasks or problems. We can find which ten of these will serve best as a test of the ability measured by the entire hundred, by measuring a suitable group in respect to the entire hundred, and choosing that set of ten whose combined score correlates most closely with the combined score for the hundred (the pairing being by individual pupils). This method not only demands much time in experimentation and calculation, but also a thorough understanding of the general logic and technique of measuring relations, including the peculiar relation of resemblances or mutual implication. For example, suppose that twenty children, who had in a given year been absent, respectively, 100 days, 95 days, 90 days, etc., down to 5 days, attained, respectively, average academic grades of 74, 75, 77, 77¼, 77½, 78, 78¼, 78½, 79, 79¼, 79¾, 80, 80½, 80¾, 81, 81½, 81¾, 82, 83 and 85. The correlation by any of the stock methods comes out as 1.00, and the investigator might fancy that he had proved that for the school in question the mark to be received could be perfectly prophesied from, and was determined by the same causes as, the attendance record. Nothing of the sort, I assure you, is proved by this perfect correlation. It would perhaps puzzle some of us to tell just why.

The method of observing the correspondence of an individual's status in the trait

[2] Average of three series.

[3] Or nearly all.

taken as a symptom with his status in other traits of whose condition it may be a symptom is of course of wider application than shown in the case supposed. An individual's status in anything may be thus compared with his status in anything else and the correspondence measured. Achievement in school studies may be compared with achievement in adult life; how far skill in motor control and craftsmanship is a symptom of ability in abstract intellectual operations may be determined; whether mental ability is or is not indicative of mental health may be decided; what an individual's interests mean for his capacities, or what his ability in any one test means concerning his ability in any other, may be found out.

One special application of the method of correlation is so important as to deserve separate mention. This is its use in testing a symptom by its correlation with some future condition. When a psychological or educational measurement of an individual enables us to make a successful prophecy concerning the individual's future and one that could not otherwise have been made, we have not only the most impressive but also the soundest proof of the significance of the symptom in question. Thus we rightly increase our faith in Mr. Courtis's analytic tests of arithmetical abilities, if we find that he is able to prophesy truly that a certain student's work in algebra will be greatly improved by certain specified drills in the fundamental operations in arithmetic, though the usual observations of her teachers gave no hint of this.

The most important accomplishment of the last decade's study of intellectual and moral diagnosis has been to establish principles of method for testing symptoms. But there has been also a substantial beginning in accumulating facts of symptomatology which education and the other social arts can henceforth use. These I shall not try to summarize, but only to illustrate.

As our first illustration of the knowledge of intellectual diagnosis that has been attained and may be expected soon to be much increased, let us take the case of general intellect—the ability to manage ideas —the average station of a man in respect to that group of powers which we call intellectual.

In order to avoid technicalities and to gain clearness I shall go somewhat beyond the specific results obtained by Burt, Bonser, Norsworthy, Simpson, Spearman, Krueger, Peterson and myself in testing tests of general intellectual ability. I shall in fact be guilty of prophesying what would occur rather than measuring what has actually been found to occur. However, the prophecy follows by straightforward inferences from facts found by one or another of these workers. Moreover, since I shall be careful to underestimate rather than overestimate the closeness of the correlation between an individual's general intellectual ability and his ability in any of the tests, the prophecy will be safe as an argument that the results of the study of mental tests are beginning to be important for education and the other arts of social control of human nature.

Suppose the men and women of this audience were measured in respect to these eight tests, four trials of each being given:

1. Supplying words to make sense in mutilated passages, the four trials being of four grades of difficulty.

2. Giving the opposites of words, each trial comprising twenty words, the four trials being of four grades of difficulty.

3. Memorizing a given word in connection with a given form, as in Fig. 1, so as to give the former when the latter is presented (there being ten pairs in each "trial").

4. Selecting from 50 forms, as in Fig. 2, 25 forms previously shown and examined for a minute or two. The 25 forms are shown in Fig. 3.

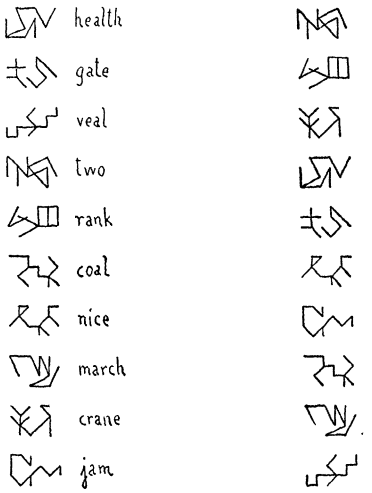

Fig. 1

5. Marking the necessarily false statements in mixed series of false and true statements, the four trials being of four grades of difficulty.

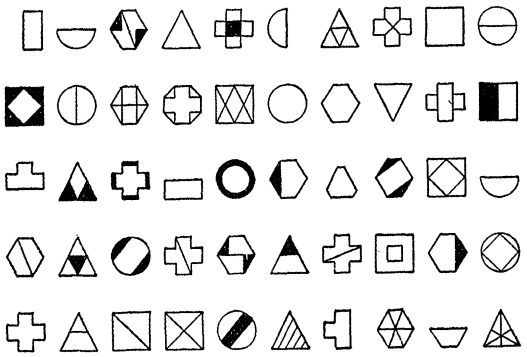

Fig. 2

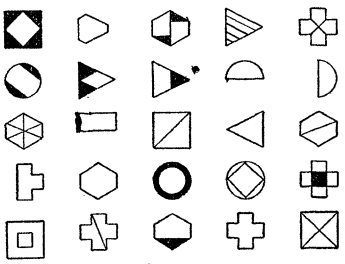

Fig. 3

6. Addition.

7. Doing what is directed in such instruction-sheets as:

Cross out the smallest dot . . .
How many ears has a cat
Make a line across this line |
etc., etc.

and

With your pencil make a dot over any one of these letters: F G H I J and a comma after the longest of these three words: boy mother girl. Then if Christmas comes in March, make a cross right here ..... but if not, pass along to the next question and tell where the sun rises ..... If you believe that Edison discovered America cross out what you just wrote, etc., etc.

8. Selecting valid from invalid reasons for a given fact, the four tests being of four grades of difficulty.

The time required would be approximately two hours, say thirty minutes a day on four days chosen at random.

From the combined score made by an individual in these eight tests, his general intellectual ability—his capacity, that is, for science, scholarship and the management of ideas of all sorts—could be prophesied with a surprisingly small error. Suppose that the general intellectual ability of the dullest men who are able to look after and support themselves (men who though temperate and strong earn say $400 a year in good times in New York City) be represented by a and that of Aristotle or Goethe by $a+b$, the difference, b, being 100. Then the amount of such ability assigned by the test alone would not, on the average, vary from the individual's true amount by more than 5; and would not vary therefrom by more than 14 in one case out of a hundred. The 5 and 14 are very cautious estimates, 4 and 11 being probably nearer what such an experiment would in fact reveal.

If each of us knew all in the company well and wrote down the names in an order of general intellectual ability, and if all of these orders were combined into an order representing the impartial judgment of the

total group about each of its members, this order would be hardly any truer than the order got by using the combined scores of the tests alone. The two orders would indeed be practically identical. There is excellent reason to believe that it is literally true that the result of two hours' tests properly chosen from those already tested gives a better diagnosis of an educated adult's general intellectual ability than the result of the judgments of two teachers or friends who have observed him in the ordinary course of life each for a thousand hours.

There might, of course, be amongst this group certain individuals of great but very highly specialized intellectual ability, as in music or military strategy or mathematics, who would be rated much lower by this team of tests than by the general test of life itself. There might also be certain individuals especially able in just the lines of intellect which these eight tests measure, but low in others, so that their rating by these tests would be much too high. However, such extremely specialized intellects seem to be rare, and this team of tests covers a wider range of the various factors to which men assign credit as intellectual ability than a superficial examination reveals. No student of the matter would pretend to diagnose a man's station amongst other men for total intellect perfectly in two hours, or in twenty! But we can by these tests measure approximately the general intellectual ability of educated men when personal knowledge of them and their achievements is lacking, and can enrich and improve on that knowledge, when it is present.

As a second illustration I choose the very practical case of prophesying how long a pupil will continue in high school.

Dr. Van Denburg secured certain information from each of a thousand pupils entering the public high schools of New York City in February, 1906, and then kept track of the length that each continued in the high school. Consider first the significance of the pupils' answers to the following questions, obtainable in ten minutes or so on the very day that a pupil enters the high school: "What do you intend to do for a living?" "Do you intend to stay in high school four years?"

If we let a line two inches long represent eight terms or four years, and draw below it the lines representing the median length of stay[4] in high school of each of the groups of pupils defined by one sort of answer, we get the facts shown in Fig. 4.

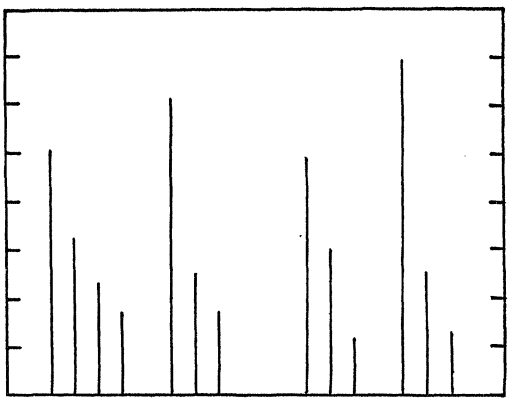

FIG. 4

It is the case, of course, that the prophecy made for any individual is only one of probability. A pupil who reports herself as intending to complete the course may stay only a half year and a pupil who reports himself as intending not to complete the course may remain for four years. Only in the long run and on the average can we be certain that the latter will stay five times as long as the former. The expectation for the two groups is shown more

[4] That is, the length of stay that half of the pupils in the group in question stay longer and half not so long.

completely in Fig. 5, where the area over each successive fraction of the four years represents the percentage of pupils who drop out from high school during that frac-

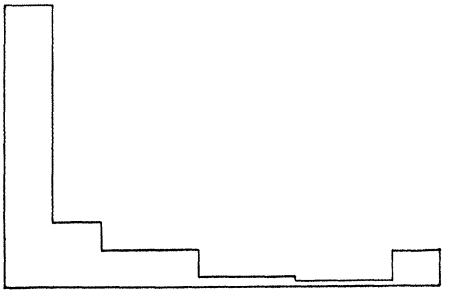

FIG. 5a

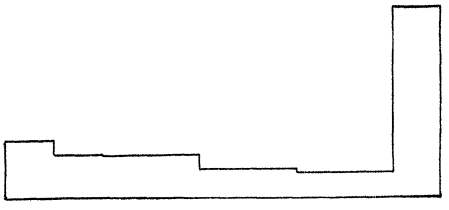

FIG. 5b

tion of the four years. Thus the chance that the "no" pupil will drop out before the second half year is 58 out of 100, or almost five times the chance that a "yes" pupil will. The chance that a "no" pupil will stay four years is 7.3 out of 100 or less than one fifth the chance that a "yes" pupil will.

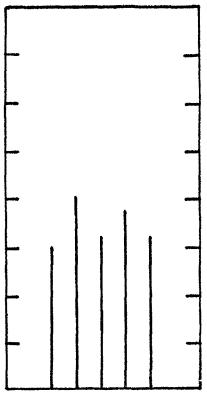

FIG. 6

An important negative fact in diagnosing the probable length of a pupil's stay in the New York City public high schools is shown in Fig. 6. Economic conditions, as measured by the family expense for rental, are seen to be insignificant.

In from one to three months after these pupils entered school they were rated by their teachers for ability, defined as follows: "Native ability apart from success or failure in any particular subject of

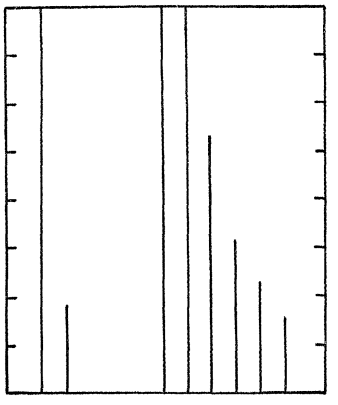

FIG. 7

study. Natural brightness." The median expectation of the top tenth and bottom tenth by the rating are as shown in Fig. 7, the detailed facts being given in Fig. 8.

For those of the pupils who stayed through the first term or nearly to its close so as to receive formal school marks in their studies, the relation of school achievement in the first term to length of stay was calculated. It is shown in part in Fig. 7 and Table II. Ten times as many of those marked below 50 leave in the first year as of those marked 90 or above. Of 115 pupils marked below 50 not one remained to graduate in four years. As the marks rise the percentage leaving in the early years steadily falls, and the percentage graduating, rises.

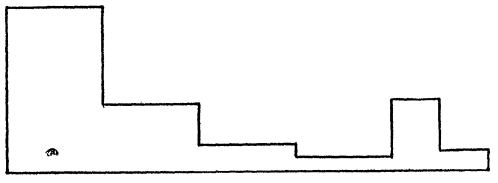

FIG. 8a

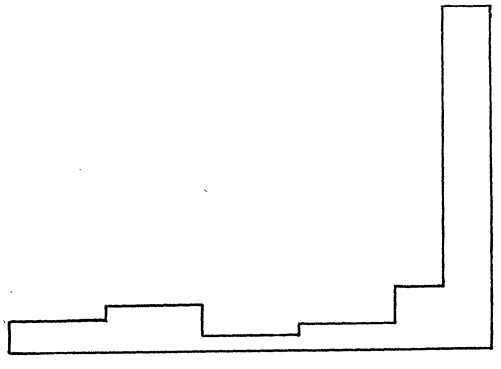

FIG. 8b

TABLE II

The relation of First Term's Mark to Length of Stay in High School

First Term's Average Mark	Percentages Leaving School Each Year After Entrance					
	Left During First Year	Left During Second Year	Left During Third Year	Left During Fourth Year	Remained Four Years Without Graduating	Graduated
Below 50	61	22	8	4	6	0
50 to 59	49					5
60 to 69	39					3
70 to 79	20					18
80 to 89	17					37
90 to 100	6	12	12	6	12	54

Such prophecies as these for New York City could easily be worked out for any community. They show that in the important matter of length of stay in school a pupil's career is far from being a matter of unpredictable fortuity. Useful diagnosis of him and prognosis concerning his high school career can begin before he sets foot in the school, and for that matter, as other facts could be adduced to show, before he is born.

These two specimens of recent work in educational diagnosis give a just notion of what may be expected from the scientific study of human capacities, interests and careers. The worker who will make repeated tests of the general intellectual development or the special school achievement of any hundred human beings for any five years of their lives, utilizing the logic and technique appropriate to mental and social measurements, may be sure of contributing to the advancement of educational science.

The work of observing and measuring the relations between traits of man's constitution, circumstances of his environment and events in his career, both as these happen in nature and as they are modified by ingenious experiment, is sure to increase our knowledge of his nature and our power over his fate. It will not be long before the members of this section will remember with amusement the time when education waited for the expensive test of actual trial to tell how well a boy or girl would succeed with a given trade, with the work of college and professional school, or with the general task of leading a decent, law-abiding, humane life.

This work of testing tests—of measuring the relations of this and that feature of a man's educational life-history—has been neglected by many of the ablest students of human nature and education, partly, I think, because it seems to lack the inspiration of sweeping theories and the drama of immediate consequences. From the ordinary point of view it is a little trivial and tedious. But in the sense that the law of gravitation has a grandeur far beyond that of the heavens—in the sense that a change in the death-rate is the most truly dramatic event in nature—in this sense the task of testing tests gives way to no scientific work in dignity and humaneness. Tables of cor-

relations seem dull, dry, unimpressive things beside the insights of poets and proverb-makers—but only to those who miss their meaning. In the end they will contribute tenfold more to man's mastery of himself. History records no career, war or revolution that can compare in significance with the fact that the correlation between intellect and morality is approximately .3, a fact to which perhaps a fourth of the world's progress is due. Experiments measuring the effects of school subjects and methods seem pedantic and inhuman beside the spontaneous tact and insight of the gifted teacher. But his personal work is confined by time and space to reach only a few; their results join the free common fund of science which increases the more, the more it is used, and lives forever.

E. L. Thorndike

W. G. Wright

William Greenwood Wright died on Sunday afternoon, December 1, 1912, in the eighty-third year of his age. He had been in apparently good health and spirits for some time past. He was found dead sitting in his chair, a newspaper fallen from his relaxed grasp. The cause was heart failure.

He was born near Newark, New Jersey; his early education was limited. He was a soldier in the Union army during the civil war, and soon after the close of that conflict he must have come to California, where he resided a few years in Los Angeles, where his only child was born and died in infancy. He went to San Bernardino about 1873, where he resided until his death and where he conducted a planing mill. About fifteen years ago he retired from active business, and spent his time in collecting and gathering material for his work on butterflies. His wife died a number of years ago, and he leaves no near relatives.

His collection of butterflies and library he has left to the California Academy of Sciences, San Francisco. Some other collections are to be sold. Mr. S. B. Parish, the noted botanist, and the executor of Wright's estate, has given me the few data now obtainable.

W. G. Wright traveled all over the west coast from Alaska to Mazatlan, Mexico, collecting specimens in various departments of natural history, but especially the Lepidoptera. He published an interesting account of his travels in Mexico in *Zoe*. An article in the *Overland Monthly* for 1884 is entitled "A Naturalist in the Desert," and an article on "Collecting in Alaska," which can not now be located. Other papers are found in *Entomologica Americana, Canadian Entomologist, Papilio, Entomological News* and Edwards's "Butterflies of North America." Perhaps the most important service he rendered to science was the help he gave to W. H. Edwards in the great work just mentioned. In the *Ornithologist and Oologist,* for February, 1885, we find an article on "An Experiment in Bird Taming," with *Phainopepla nitens;* his name is frequent in the two large volumes, "Botany of California," as he was an enthusiastic collector of plants. His most important book, "The Butterflies of the West Coast," was published in San Francisco in October, 1905, and was really an epoch-making publication, notwithstanding the numerous inevitable mistakes. This work was illustrated entirely by color-photography.

Among the insects which have been named in his honor by different men, are: *Melitæa wrightii, Copœodes wrightii, Gluphisia wrightii, Leptarctia wrightii* and *Selidosema wrightiarium.* He named a number of new species, but a good many of these, especially those in his 1905 book, are synonyms.

Mr. Wright was a close friend of the two noted pioneer botanists and collectors, Edward Palmer and C. C. Parry, and made many excursions, of varying lengths, with them. He knew many other botanists and entomologists also.

W. G. Wright will always be remembered by those who were so fortunate as to have known him personally. He was a *naturalist*