

Intelligence, IQ Public Opinion and Scientific Psychology

Alfred H. Shephard

University of Manitoba

Many psychologists have made unjustifiable inferences from data by interpreting IQ as a causal variable when logically it is a non-causal variable. Attempts are made in this paper to understand how the confusion came about and to suggest initial strategies for eliminating it. Binet's successful prediction of school performance from test scores did not encourage him to conclude that he had defined intelligence—both he and Boring accepted a co-variation model, but only on an interim basis. A failure to distinguish between differences in the logical structures of causal and non-causal explanations, together with Goddard's misunderstanding of Binet's view of intelligence test scores, led to a misuse as well as a misconception of IQ. After Boring calmed the public outcry over many psychologists' interpretation of World War I test results, a causal view of test score in relation to school performance was encouraged, culminating in ideas expressed by Jensen (1969) on black-white differences in IQ. These views led to such uses of IQ as assigning black children to remedial programs, an activity which some law courts in the United States have ruled discriminatory and illegal. Causal and non-causal models are differentiated on the basis of identified and unidentified sources of variability in school achievement. Speculations are made on reasons for the continuing confusion between these two models, and suggestions are offered for a more discriminating use of the two types of concepts.

Though thought to be very useful in the past for planning educational programs, IQ is now regarded by some responsible public bodies as discriminatory and illegal. "It is made all the more invidious when legitimated by ostensibly neutral scientific IQ scores," was the observation of Federal District Court Judge Robert Peckham in his decision to invalidate the California Department of Education's use of IQ for assigning black children to remedial programs for the educable mentally retarded (Peckham, 1979). Why are intelligence test scores, which both facilitated the recognition of psychology as a science and contributed greatly to its public credibility and support, now viewed with doubt by scientists and public alike?

Part of the answer to this question is that a confusion has developed between two classes of psychological variables which are logically (i.e.,

structurally) quite different. We will refer to one of these classes as *causal* and the other as *non-causal*. In the description of human behavior, *cause* has had different meanings for different scholars (Harré & Madden, 1975; Hume, 1960; Michotte, 1963; Shultz & Ravinsky, 1977). The meaning we give causal and non-causal will become clear later when we distinguish between the logical structures of the two classes. It is difficult to find a language devoid of all associative or surplus meaning, but causal and non-causal do serve at least to emphasize the crucial difference between the two. Presenting a persuasive argument is made all the more difficult because some of the language used in our analysis will inevitably be the same language that is itself inextricably enmeshed in the confusion. "We have great difficulty in representing the world of experience to ourselves without the spectacles of the old-established conceptual interpretation....our language is compelled to work with words which are inseparably connected with those primitive concepts" (Einstein, 1935).

Our main concern will be with the formal structures defining, and the mathematics representing, these two categories. *Model* will be used in referring to each category to emphasize that, in the first instance, the definition is arbitrary—a representation of nature. Thus a causal model is a variable whose definition is structured to permit causal attribution statements (e.g., level of illumination could be defined in such a way as to permit differences in visual acuity to be attributed to differences in level of illumination). A non-causal model is a variable whose definitional structure permits no causal attributions (e.g., score on a pilot aptitude test). My contention is that a failure to make the logical distinction between the operations defining these two classes of variables has led to statements which are not justified by the logical structure of the non-causal model. Of all psychological variables IQ has perhaps suffered most from this lack of discrimination between causal and non-causal variables. Possible reasons for the perpetuation of this ambiguity are suggested and means are proposed for achieving a less confused understanding of the complex domain subsumed under intelligence.

Our Starting Point

Alfred Binet, both an experimental psychologist and an experimental pedagogue, is regarded as a pioneer in intelligence testing. Yet, as Wolf (1973) remarked, "up to the time of his death Binet felt that the lack of sufficient evidence justified his avoidance of a definition of intelligence or of broad hypotheses about its nature" (p. 216). He was challenged by the practical problem that arose from a decision in 1881 to enforce universal education in France (Wolf, 1973, p. 160). Clearly, there were

extreme differences in school achievement among students. As conscientious teachers will avow, such differences may create havoc in the standard classroom teaching situation. This problem for the Paris schools was viewed as one of classifying or screening out the retarded pupils — mainly those referred to as *les débiles* (morons) — before they appeared in the regular classroom. Special classes could then be provided for them (Commission...1904). While the scientific community accounted for differences in school achievement on the basis of differences in intelligence there was no generally accepted means for distinguishing one student from another in this regard.

Binet succeeded in developing a test which proved useful in predicting school achievement. In spite of this success his puzzlement about how school achievement, test performance, and intelligence were related is expressed in the following statement:

We must make known the meaning we give to this vague and very comprehensive word "intelligence." Almost all the phenomena that occupy psychology are phenomena of intelligence; sensation, perception, are as much a part of intelligence as is reasoning. Should we therefore introduce into our tests a measure of sensation such as is used by the psychophysicists? Should we put all of psychology into the tests?

After some reflection, such an attempt appears a waste of time. There is in intelligence, it seems to us, a fundamental agent the lack or alteration of which has the greatest importance for everyday living, and that is judgment, commonly known as good sense, practical sense, initiative, the faculty of adapting oneself to life. Good judgment, good understanding and sound reasoning, these are the essential domains of intelligence. (Binet & Simon, 1905a, p. 196)

The quest for a definition of intelligence still continued after Binet's death. He worked within the general belief that some as-yet-unidentified concept, loosely referred to as intelligence, would account at least in part for differences not only in school performance but also in all other adaptive behavior. Although differences in school achievement might be regarded as the result of differences in intelligence, Binet did not agree that the definition of a variable which successfully predicted school achievement was necessarily the definition of intelligence. As we shall see his rejection of the test as a definition suggests that he demonstrated some sensitivity for a difference between causal and non-causal models. Further developments in the concept of intelligence were importantly influenced by the fact that many researchers did not display this sensitivity. An assessment of these developments over the last 75 years can be expressed in a language comparable with that used by Kendall, regarding the development of the concept of probability in statistics: "knowledge of the development of the subject would have rendered superfluous much of what has been written about it" (1956, p. 11).

Measurement and Scales

In this paper, “measurement” will mean the scientific use of numerals to represent attributes such as height, weight, and age, from which scales emerge—the familiar nominal-ordinal-interval-ratio designation. The meaning of the numerals depends on the nature of the operation for assigning them to separate instances of the attribute. As this operation is the scientists’ way of defining differences in the variable, it is generally referred to as an operational definition. Bridgman’s (1927) argument on operationism would seem to include the following two elements: that the actual label or name given to an empirical variable is of no importance so long as it is discernable from other names and all scientists agree on it, and that the operation for comparing one instance of the variable with another in a quantitative manner is explicit, clear and public.

Prior to the 1905 test, Binet and Simon were well aware that intelligence was a label without an operation and that the label not only was meaningless in the scientific operational sense, but also was rich with years of non-scientific surplus meaning. Binet insisted that whatever the 1905 test meant, it did not define intelligence. But did it, as Messick (1975, p. 957) might suggest, measure something? Binet himself had misgivings about the question of whether he was carrying out operations that could legitimately be called measurement. He apparently differed with his collaborator, Théodore Simon, on this question (Binet & Simon, 1905a, p. 194), insisting on a distinction between measurement and classification, with test scores providing classifications rather than measurements. In discussing this distinction, Binet (1898, p. 113) recognized that sensory variables could be defined by operations that permitted measurement, but the “richness of intelligence” could not be captured by such a definition. Yet, when referring to test scores he continued to confuse his readers by using terms such as “measure” and “scale” (Binet & Simon, 1908) and “precision” or “accuracy of measurement” (Binet, 1911).

Multi-sources of Variability for School Achievement and for Test Score

For Binet, intelligence was part of practically all topics of interest in psychology (Binet & Simon, 1905a, p. 196); and consequently, all behavior was a function of intelligence. However, he recognized that any behavior might also be related to numerous other variables. Measuring intelligence would necessitate the study of behavior in all possible circumstances (Binet & Simon, 1908) in order to separate it from these other variables. In any situation the behavior exhibited would be, he said, a function of four categories of variables: natural intelligence, school

learning, extra-curricular acquisitions, and language and vocabulary acquisitions. Natural intelligence would seem to be that source of variability which is due to hereditary factors. School learning would include various sets of possible sources of variability related to the quality of the school, the staff, the curriculum, teaching methods, teacher and fellow-pupil personalities in relation to the pupil's personality, and instructional aids. Extra-curricular acquisitions would include various sets of variables not specifically associated with the school such as parent education, socio-economic status of the family, hobbies, birth order, and characteristics of the child's siblings and peer group. Binet might have included language and vocabulary learning simply as part of the two experiential categories of variables, but instead he set up the fourth category which mirrored joint effects of all experiential variables on language acquisition.

Bewildered by the complex pattern of sources of variability subsumed under his four categories he asked what differences between test scores could possibly mean. He noted that the meaning of a test score was difficult to assess if a very intelligent child had been denied school instruction because he/she lived too far from school, had a prolonged illness, or was kept home for household chores, such as caring for a sick relative or herding the sheep (Binet & Simon, 1908). After studying test scores for children aged three months to two years from a day-nursery, Binet and Simon commented: "Even from this age, extreme poverty, the absence of fondling and being played with, already makes its mark and retards the awakening of intellectual faculties" (Binet & Simon, 1909, p. 4). Their deliberations led to the identification of variables that did, and conjecture about variables that could, influence the test score.

In pondering the meaning of the test score Binet was considering variables other than natural intelligence which might reasonably have accounted for differences between test scores. In our terminology, he was proposing specific variables which might have caused the differences. He was striving for causal models of test score and of school achievement as alternatives to the non-causal model linking the two. While his arguments did not include our terminology, their form was consonant with the structural constraints of the causal and non-causal models that we will shortly consider.

How far had the empirical scientific understanding of intelligence progressed up to the time of Binet's death in 1911? There certainly was not an agreed-on operational definition. Binet did not regard the test as an operation defining intelligence. The test was validated simply by noting that values of test scores, differing in an unexplained way, tended to covary with corresponding differences in levels of school achievement, similarly unexplained. Level of achievement in school could be predicted

with some precision from test score, even though the sources of variability of neither level of school achievement nor of test score were known. Since no sources of variability are known, we will refer to research of this paradigm as a "non-causal model."

From Binet to Terman: Trans-Atlantic Confusion

Following the 1908 revision of Binet and Simon's test, there was a general interest in their work, particularly in the United States where Goddard proselytized with great vigor. "Yet it often happens that the devoted disciple transforms the ideas of the prophet in the very process of transmitting them. So it was in this [Goddard's] case" (Tuddenham, 1966, p. 490). Goddard enthusiastically employed and publicized the test but misconstrued Binet's views of intelligence (Goddard, 1916). For Binet, intelligence remained undefined and was vaguely thought of as pervading all behavior, with its importance and significance relative to other sources of variability assessed for each behavior—empirically, an almost impossible task. For Goddard (1920), differences in school achievement were due almost entirely to differences in intelligence, which he regarded as a unitary variable resulting largely from differences in heredity, therefore fixed and immutable—a view much at variance with Binet's.

With the transplanting of test-intelligence to America a sequence of events began that served to move psychological thinking concerning intelligence away from Binet's views. For Binet, "intelligence" was an undefined word; "test score" was a useful means of predicting school achievement. For Goddard, "intelligence" was a hereditary variable responsible for feeble-mindedness; "test score" was intelligence. The feeble-minded child required institutional care different from that provided by the regular school system—a problem not much different from Binet's problem of providing special classes for *les débiles*, but conceptualized in a logically quite different way.

Confusion Perpetuated

A development that further served to move the consideration of intelligence away from the empirical, multi-source approach preferred by Binet was the fact that, "by default, Spearman's theory came to constitute the conceptual basis for Binet's test approach" (Tuddenham, 1966, p. 504). Binet's failure to provide either a definition or a theory for intelligence was simply the result of his conviction that sufficient research had not been done to permit psychology to advance in this way. He opted, for the time being, for a straight empirical approach, both in

predicting school achievement and in advancing an understanding of intelligence. Wolf (1973, p. 216) remarks that, as Binet (1909, pp. 242-243) had openly not accepted Spearman's (1904a; 1904b) formulation, it was ironic that this formulation should subsequently be allotted the role of supplying a theoretical framework for Binet's empirical research.

Problems concerning the nature, definition, and views of intelligence which interested Binet and early American psychologists were even further removed from center stage by arguments over the meaning of World War I testing and the manner in which scientists interpreted Boring's article (1923) on intelligence test scores. Three years after the termination of World War I, psychologists were still asking (Intelligence, 1921) the meaning of scores from a test they had administered extensively during the War. The embarrassment and humiliation that almost certainly would have resulted from fictions (Pastore, 1978) perpetrated on the public was ingeniously obviated by Boring's statement that "intelligence is simply what the tests of intelligence test" (1923, p. 35). Boring had turned a possible demise into a victory and a profession. Regardless of interpretation, his idea of an operational definition allowed psychology to proceed as an empirical science rather than to become enmeshed in philosophical debate regarding the definition of intelligence. Predictions could be made with some precision, providing a scientific answer to a social problem.

Although Boring's paper tended to move scientific psychology still further away from Binet's hopes for a causal understanding of behavior in general and of intelligence in particular, Boring's intention was that his paper should be regarded as an interim understanding of an unfortunate *fait accompli* until, as Binet had suggested, more empirical research could be undertaken:

It would be better if the psychologists could have used some other and more technical term, since the ordinary connotation of intelligence is much broader. *The damage is done*, however, and no harm need result if we but remember that measurable intelligence is simply what the tests of intelligence test, until further scientific observation allows us to extend the definition [*italics mine*]. (Boring, 1923, p. 35)

In fact Boring's quick-witted interim understanding rapidly became misinterpreted as a final causal solution (i.e., differences in school achievement were viewed as being caused by differences in intelligence).

While psychologists were relieved that the bitter public debate had been eased by Boring, any misgivings about the scientific acceptability of such a solution were eliminated by another development in the history of science. Physics, that pioneer for disciplines seeking a scientific understanding of their domains, had problems regarding the status of empirical concepts arising from aspects of Einstein's restricted and

general theories of relativity. Bridgman (1927) notes, "it was a great shock to discover that classical concepts, accepted unquestioningly, were inadequate to meet the actual situation" (p. 1). However, in a manner not unlike Boring's proposal for the meaning of a test score, Bridgman advanced an understanding of the general structure or logic of the definition of a scientific empirical variable, at least for physics: "In general, we mean by any concept nothing more than a set of operations; the *concept is synonymous with the corresponding set of operations*" (Bridgman, 1927, p. 5). As noted previously, the nature, definition and measurement of intelligence had been a profound challenge for psychology. But as the "great shock" in physics quietly subsided, Boring's cavalier-appearing comment on intelligence seemed justified.

Just as the success of the Stanford-Binet test (Terman, 1916), serving as an operational definition for intelligence, freed psychologists from worrying about the puzzling problem of intelligence and gave them plenty of time to construct many other tests for various purposes, so Bridgman saw a general advantage in operationism:

Possibly after everyone has schooled himself to this better way [operational thinking], there will remain a permanent unsocial tendency, because doubtless much of our present conversation will then become unnecessary. The socially optimistic may venture to hope, however, that the ultimate effect will be to release one's energies for more stimulating and interesting interchange of ideas. (Bridgman, 1927, p. 32)

Undoubtedly this strategy did tend to lessen both professional and public philosophical debate. Operationism might have provided a reasonable non-causal solution to the question of the scientific definition of intelligence—an agreement that operations associated with the Stanford-Binet test, for example, would be the accepted scientific definition. Unfortunately, this was not the way methodology subsequently developed. Additional tests were constructed so that operations defining the Stanford-Binet were no longer regarded as *the* definition. Encouraged by arguments such as those concerning construct validity (Cronbach & Meehl, 1955), several tests came to be referred to as *measures of intelligence*—a bewildering development for a methodology aspiring to scientific status. The several different definitions were not justified on the basis of the closeness of correspondence of the numbers on different tests, as a physicist might do for length with a meter stick or by triangulation. Rather they were justified merely on the basis of a correlation coefficient indicating whether there was *any* correspondence of numbers at all from test to test.

Jensen's question (1969) about raising scholastic achievement and boosting IQ will be used later in this paper as an illustration of how far psychologists have deviated from Binet. It also indicates that Boring's caveat (1923, p. 35) about the "harm" that would result from lifting in-

telligence from everyday language into the precise world of science was not remembered; the interim understanding had become misinterpreted as a final causal solution of the problem of the nature of intelligence. Boring, like Binet, was undoubtedly concerned about the adequacy of the test-score understanding. One concern was the surplus meaning (i.e., much broader than that contained in the definition) that would accrue with a common parlance label. Another matter that appears to have bothered both Binet and Boring is just what "correlation" meant. Boring (p. 35) states,

If the correlation is considerable, yet not perfect, say 60 percent, we say that the particular capacity is partly dependent upon intelligence and partly independent of it. We shall not be far wrong if we think of such a capacity as complex, involving 60 percent of intelligence and 40 percent of some special ability that is not intelligence.

Clearly Boring, like Binet, would appear to be striving for what we can best describe as a causal analysis of behavior (i.e., he wanted to identify intelligence as *the cause of 60%* of the differences in school achievement, and non-intelligence as the cause of the remaining 40%). As will be noted shortly, the logical structure of only the causal empirical representation permits such an attribution. As previously noted, intelligence defined as a score on an intelligence test is a non-causal representation.

Cronbach (1957; 1975a) has also discussed concepts we have designated causal and non-causal. In this paper we will point out differences in their logical structures and in their appropriate mathematical representations to highlight characteristic differences in empirical statements in which they might legitimately appear. As with probability and empirical events, both causal and non-causal concepts are scientific ways of representing nature—names for empirical observations (Shephard, 1978).

The causal model emerged earlier than the non-causal model. Though measurement is an essential aspect of empirical science, values of variables represented by numerals are subject to variability—among observers and among observations for the same observer. Some way was sought for determining a precise value of a variable under these conditions. A suitable method was found by constructing a model to present the fact of variability in observations.

Causal, Gaussian, or Variance Model

Boring (1950, p. 134) relates the story of how Maskelyne, the Astronomer Royal, in 1796 at the Greenwich Observatory, dismissed his assistant, Kennebrook, because there was a discrepancy of 0.8 sec in their observations of stellar events (recorded time of the star crossing crosshairs in the telescope) and also of how Bessel came to study the per-

sonal equation or variability in empirical observations. Since safety in maritime navigation would depend on accuracy, this question was of paramount importance. While there was variability in times recorded, values seemed to cluster within a relatively small range. Which value was true, correct, or accurate? There was no way to select one over another.

Scientists were agreed that the magnitude of each observation was related to a number of possible sources of variability, at least one of which was the same for all observers: the hypothetical "true" time. Other variables that might differ among individuals or within the same individual from time to time might be fatigue, genetic factors, health variations, and anxieties regarding family income. All possible sources of variability were divided into two classes, the identified component (stellar crossing) and an unidentified component. (Other terms could also be used to characterize these: specified and non-specified, constant and variable, homogeneous and heterogenous, and known and unknown). The model, for obvious reasons, became known as the theory of errors and was examined by Simpson (1757) and later by Lagrange (1770), who developed what was referred to as an error curve.

Unfortunately, even after all this modelling, the observational error was no different from what it was originally—the fallible measure was still equally fallible—but was now interpreted in a way that recognized the condition of a constant value of a variable for each value of an aggregate of astronomical observations. The combined effects of unidentified and unidentified components were represented by a central value. The potency of the unidentified component alone was represented by the extent of values on either side of it. The Gaussian components were typically represented as follows:

$$\bar{X} = \sum_{i=1}^n X_i/n \qquad SS = \sum_{i=1}^n (X_i - \bar{X})^2$$

\bar{X} , mean representing the combined effects of identified and unidentified components,

X_i , value of any observation,

n , number of observations,

SS , sum of squared deviations representing potency of the unidentified component alone for n observations.

For any set of n observations, neither the relative contributions of identified and unidentified components to the value of \bar{X} nor the contributions of the unidentified component alone to SS is known.

In the typical experimental study, a possible source of variability is identified by changing its value (for example, from one group to another). The researcher must decide whether differences between the groups can be attributed to differences in the identified variable. One possible source of variability is known—the identified. Mathematical models have been developed to allow a comparison of the combined effects of the identified and unidentified components with unidentified alone.

Our interest in this paper is not in the manner in which the scientist uses random probability models to deal with differences from unidentified sources. More important is the fact that the variables of the mathematical models correspond to the representations of the two classes of possible empirical sources of variability, \bar{X} and SS , noted above.

The following Student's t distribution is a probability model frequently used in causal analyses to attribute at least some of the total variability of empirical observations to identified sources, notwithstanding a possible unidentified component:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{SS_1 + SS_2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \cdot \quad df = n_1 + n_2 - 2$$

\bar{X}_1, SS_1, n_1 mean, sum of squared deviations, and number of values for one value of the identified variable,

\bar{X}_2, SS_2, n_2 mean, sum of squared deviations, and number of values for a different value of the identified variable.

In this null hypothesis form it is conjectured that the model best representing the research data is one in which the identified variable is not relevant. On the basis of the value calculated for the model, a decision to reject or fail to reject the hypothesis is made. Given this hypothesis, any difference between \bar{X}_1 and \bar{X}_2 or SS_1 and SS_2 would logically be due to unidentified sources. A decision to reject the hypothesis results in a statement that some of the total variability can be attributed to the manipulated or identified variable. It is in this sense of knowing one possible source of variability that we will refer to the identified variable as permitting an *attribution* or causal statement.

Galton's Preservation of Error

While Gauss was mainly concerned with the application of what was referred to as the normal law of error, "his name became attached to it in the way that promoters are often confused with discoverers in the history of science" (Boring, 1950, p. 499). Quetelet (1835) found many empirical situations that could be represented by the Gaussian model, and Galton's enthusiasm for it can be gathered from the following:

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error." The law would have been personified by the Greeks and deified, if they had known of it. (Galton, 1889, p. 66)

Yet, as we shall see, he subsequently rejected it as being unsuitable for his research.

In his work on heredity and its relation to various attributes, Galton felt that if he could show that offsprings were more like their own parents than any other parents, he would have a persuasive argument for stating that, "family likeness and individual variations are largely due to a common cause"—a "factor of stability" (Galton, 1889, pp. 11; 196). He wished to designate this common cause, "heredity." However, as is frequently the case with empirical observations, the data failed to show a clear one-to-one correspondence of offsprings with parents: there was variability in correspondence of progeny and parent. As with the stellar transit problem, where variability also characterized the observations, some model was sought to represent closeness of correspondence among offsprings and parents. Karl Pearson (1924, p. 392) indicated that one of the earliest records of a bivariate comparison was Galton's unpublished tabular representation, (appearing about 1875), concerning size of parent and offspring sweetpeas, with both regression lines drawn.

Although the table was, as Pearson suggests, a first attempt at a representation of covariation of the two sets of values of variables, it was a radical departure from the powerful mathematical Gaussian model of which Galton was so enamored. Galton said he could readily comprehend the bivariate tabular representation, "but I could not see my way to express the results of the complete table in a single formula" (Galton, 1908, p. 302). Pouring over the diagram in his notebook, he noticed that the lines of equal frequency formed concentric ellipses, and he referred this characteristic of the data to a mathematician to see whether progress toward a goal of a "single formula" could be developed further: "I did so, under the form of a problem in mechanics, and he most cordially helped me by working it out, as proposed, on the basis of the usually accepted and generally justifiable Gaussian Law of Error" (Galton, 1908, p. 303). The solution was included as an appendix to Galton's paper (1886), but

since it depended on the relative size of the empirical measuring units, the search continued for a unit-free model.

Galton's struggle to find a suitable model for his research data clearly illustrates not only the difficulty that may be experienced in any attempt to construct a mathematical model but also how easily an otherwise excellent model can be misused. The question for Galton was not whether there should be a model but what aspects of the empirical situation were to be represented in it. He expressed his appreciation for the assistance that his mathematical friends had provided, and added:

The patience of some of my mathematical friends was tried in endeavouring to explain what I myself saw very clearly as a geometrical problem, but could not express in the analytical forms to which they were accustomed, *and which they persisted in misapplying* [italics mine]. (Galton, 1908, p. 305).

He mentions, for example, advice he had received regarding the use of the Gaussian model for the general paradigm of research that he was conducting. He comments:

Still, he was quite wrong. The primary objects of the Gaussian Law of Error were exactly opposed, in one sense, to those to which I had applied them. They were to get rid of, or to provide a just allowance for errors. *But these errors or deviations were the very things I wanted to preserve and to know about* [italics mine]. (Galton, 1908, p.305)

Cronbach (1957, p. 671) makes the same point somewhat differently: "While the experimenter is interested only in the variation he himself creates, the correlator finds his interest in the already existing variation between individuals, social groups, and species."

Non-causal, Galton, or Covariance Model

Galton's justification for rejecting the Gaussian model for his research clearly suggested the general form of a suitable model. Heredity was an obvious variable to consider in relation to understanding likenesses among parents and offsprings, because without heredity there would be no individual. For the layperson, it was clear that there were among-families and among-family-members differences in heredity, and these frequently were cited as sources of readily discernable physical and psychological differences. But the task of making scientific observations of human variability that permit the isolation, identification, and attribution of heredity as a source of this variability appears difficult, even at present with a causal model. As we shall see, this is logically impossible within the formal structure of a non-causal model. Galton (1888) first used the term *co-relation* to speak of likeness among parents and offsprings, giving the term *correlation* a unique non-causal implication. Both parents and their offsprings showed differences, but no sources of

variability were identified for either. Values for each child were simply yoked with values for their parents. While the causal model represents two components of variability, one identified and one unidentified, the non-causal model represents two components, neither of which are identified—*both* are unidentified.

The following Pearson product-moment correlation model—representing, for example, the extent to which covariation between IQ and school achievement can be represented by a straight line—well illustrates how the effect of a single value of a variable (that could have been identified) is removed, leaving only unidentified components:

$$r_{x \cdot y} = \frac{\sum_{i=1}^N (X_i - \bar{X})}{N \sqrt{\frac{SS_x}{N}}} \cdot \frac{(Y_i - \bar{Y})}{\sqrt{\frac{SS_y}{N}}}$$

X_i, \bar{X}, SS_x , intelligence test score, mean and sum of squares,

Y_i, \bar{Y}, SS_y , school achievement level, mean and sum of squares,

N , number of pairs of observations.

By representing each of the paired values of the variables by deviations from their respective means divided by their respective standard deviations ($\sqrt{SS/N}$), the model is rendered free from measuring unit constraint, a question that had been of great concern to Galton. Also, the amount of the deviation is expressed in terms of a measure of the relative potency of the unidentified component for test scores (SS_x), and a measure of the relative potency of the unidentified component for school achievement (SS_y) for the N pairs of observations.

Some Differences between Causal and Non-causal Variables

Our main concern is with the structural differences between the two models; simply stated, causal identifies a cause and non-causal does not. For example, a causal account of variability in school achievement would note a variable such as parental interest (Heslip, 1974), previously identified, manipulated, and found relevant. A non-causal account would not identify any source of variability for either intelligence test score or school achievement but would simply use their previously determined covariation.

Nevertheless, an appreciation of non-structural differences between causal and non-causal is important for an understanding of the ease with which the two models can be confused. In comparing them, "criterion variable" (e.g., school achievement) will be used to refer to both the causal dependent variable and the non-causal validating criterion—the criterion for the researcher's decision with regard to a relation between the variables. The identified variable in causal research is initially defined without reference to the criterion variable by the operations permitting its change in value. It remains a variable regardless of whether or not the researcher decides it is relevant. In contrast, in initial non-causal research the test score is only a classification test variable if it covaries with the criterion variable. If the researcher fails to decide they covary, the consequence is a decision that the items, instructions, and scoring, do not constitute operations defining anything. Some facets of the operations are then revised until a decision of covariance can be made, thus giving birth to a classification test variable.

Another non-structural difference between causal and non-causal research is related to the strategy in causal research which divides all possible sources of variability into two classes—identified and unidentified. The random models and hypothesis testing in causal research become essentially a means of deciding whether (considering the combined potency of the identified and unidentified components relative to that of the unidentified component alone) any of the total variability can be attributed to differences in the identified variable. An attribution or causal statement is made in spite of the uncertainty of the effects of the unidentified component.

The decision to pursue a non-causal empirical understanding of the criterion variable, using intelligence or ability test scores, is also a decision not to identify any variables related either to the criterion variable or to the test variable. Undoubtedly, numerous variables could be identified to account for differences in test scores but this would be causal research. Similar causal research could be pursued for school achievement (Heslip, 1974). The non-causal empirical researcher is typically not engaged in either of these endeavors. Concern centers on predicting the criterion from the test variable, regardless of the "causes" of each. Nevertheless, even though the researcher may not be concerned with sources of variability, "already existing variation" is essential for covariability. Any misguided attempt to identify sources of variability in the test score tends to reduce both its original variability and the covariation that previously justified test score as a non-causal classification variable. Inappropriately analyzing test score variability with a causal model destroys its justification as a classification test variable. Loosely speaking, we need the intact, "already existing" variability that provided the prerequisite

covariation. Test scores cannot be analyzed and still remain classification test variables.

A Causal Concern Resigns to Non-causal Research

Binet could have persisted in a causal approach to understanding the general area he thought should be subsumed under intelligence; but he was successful with a non-causal approach, in which no sources were identified. He proposed, "for now, let us content ourselves with evaluating their intelligence in general" (Binet & Simon, 1905a, p. 193), so that an inability or unwillingness to agree on a definition for intelligence would not hinder progress in predicting level of school achievement. (Care must be taken to distinguish "intelligence in general" from a factor of general intelligence, *g*). Although scientists in the young science of psychology might not know why (in the sense of being able to identify causes) children differed in test scores and in school achievement, the non-causal test score understanding would "for now" permit the new school system to cope with wide and challenging differences in school achievement. This would seem to be the interim understanding that Binet proposed.

Binet and Simon (1905b) expressed concern about what we have called the "unidentified sources" characteristic of the test score. In speaking favorably of the Bin-Damaye test, the immediate precursor of the 1905 test, they expressed a desire to identify what we have termed causal variables related to school achievement. They preferred this to having a heterogeneous mixture of unknown sources affecting the responses to a variety of test items. What, they asked, could a total score obtained by adding up credits for appropriate responses possibly mean? A score of 70, they argued, does not always have the same meaning in the sense of getting the same items correct or of sampling the same attributes. For example, one score of 70 might be related, in part, to middle socio-economic status, a history of punitive behavior control, being a male only child, and having a white skin. Another score of 70 might be related to high socio-economic status, supportive behavior control, being a male child following a female child eight years older, and having a black skin. They were well aware of the numerous possible sources of variability in school achievement and in test score that might be identified and analyzed according to a causal model, but chose rather, "for now," to leave them unidentified as in a non-causal model.

Boosting IQ – An Indicator of Confusion

A "vast literature" based on the implicit assumption that IQ could be

equivocated with an identified variable, permitted Jensen (1969) to speak of "boosting IQ." No part of the operation defining IQ is a manipulative operation as might characterize a variable such as level of illumination. The implication of the differences between the experimenter (who "creates" variability) and the tester (who is interested in "already existing variation") is that the experimenter identifies possible causes of variability, while the tester works with unidentified variability. This lack of empirically identified sources need not deter conjecture and theory from being developed in various ways, *but none of these are substitutes of analyses using empirically identified variables.*

Undoubtedly some source of test score variability could be identified and manipulated according to the causal model to "boost" the test score. Without becoming involved in the details of the argument, suffice it to note that these new scores would not be IQ. They would not represent, as Cronbach noted, "already existing variation," but variation due partly to unidentified and partly to identified sources—a mixture of both models producing an unpalatable, uninterpretable potion.

The inappropriateness of analyzing IQ by a sampling model similar to Student's t distribution described above is more readily understood by considering the "vast literature" (Jensen, 1969, p. 81) which attempted to identify a black-white skin variable as a source of test score variability. Previously we noted that a decision regarding the effect of an identified variable could be made by using a random probability model to compare the combined effects of the identified and unidentified classes of sources with the unidentified alone. When level of illumination is the identified variable and visual acuity the criterion variable the interpretation of a rejection of the null model is straightforward—the cause of the change in visual acuity is the change in level of illumination. Black-white skin-color differs from level of illumination in that the researcher studying differences in skin-color did not assign a skin color to each of two randomly selected samples. Rather random samples were selected from each of two already established groups differentiable on the basis of color of skin. One systematic difference in this *complex variable* is known, but the method of determining the samples is such that it is not possible to state that the unidentified class of sources is random. The danger in sampling from already established groups is, of course, to assume that they differ systematically with regard to no other variable than the one identified.

Should the t -model be regarded unjustifiably as suitable and should a statement that skin color is related to test score be regarded erroneously as tenable from the outcome of a statistical test of the null hypothesis, great caution would be required in making attribution statements regarding the skin-color variable and IQ. Clearly there are possible differences related to IQ such as heredity, socio-economic status, school opportunity

and family training. Even though color-of-skin is recognized as a complex variable embracing more identifiable, systematic sources than might appear from its name, only the numerator of this inappropriately used t -model is adequate to represent the empirical research situation.

Jensen (1969, p. 81) notes that the "literature" indicates that blacks are consistently 15 points lower than whites. Our discussion thus far suggests that numerous sources, including hereditary ones, might singly or in combination account for this difference if they were identified. However, the denominator of the t -model renders this model totally unacceptable for examining black-white differences in IQ. The two main components of the denominator are the sums of the squared deviations for blacks (SS_b) and for whites (SS_w), mirroring potency of the unidentified component uncontaminated by the effects of the identified component. In the sampling model the unidentified sources are said to be "random" and "independent" of the identified sources represented in the numerator. That is, components of the unidentified sources do not tend systematically to favor the value of the numerator one way or the other, nor does the denominator show different potency of unidentified sources for different values of the identified variable—they are independent.

Although this paper is not intended as a critique of the manner in which a researcher might attempt empirically to approximate the random constraint of the model, it is obvious that in the black-white comparison of values of IQ many sources could not be represented by a random model. What these are cannot readily be decided, because partitioning of this sort would partly depend on what sources were included in the complex black-white skin-color variable considered above. However, Jensen's statement that the variability of blacks' scores is consistently 60% of whites', obviates the need to consider the unidentified component or to work out details of this question here. Clearly and consistently, according to Jensen, the potency of unidentified sources for test score is not the same for blacks as for whites and, therefore, cannot be represented by a random model. Thus it is not appropriate, using random sampling models, even recognizing skin-color as a complex variable, to compare the combined effects of skin-color and random unidentified sources in the numerator with unidentified sources uncomplicated by skin-color in the denominator—the two classes of sources are not independent. The causal analysis is not appropriate because of a failure of the empirical situation to meet the random and independent requirements of this sampling model.

A further objection to a misguided analysis of non-causal scores with a causal model is that identification of part of the previously unidentified component of test score variability would, as noted above, make covariation untenable (i.e., disrupt the "already existing variability"). Hence

justification for the classification test variable would be destroyed. The logic of non-causal research is premised on *intact unidentified sources only*.

Logical Structure and the Misuse of Variables

Lacking our modern understanding of models representing nature (Wolf, 1973, p. 34), Binet did not clearly and unequivocally set out the distinction between what we have termed causal and non-causal variables, but as Varon (1936, p. 39) has noted, Binet did distinguish "two great problems." Phrases such as "one causes a variation" (Binet & Henri, 1895, p. 422) suggest the causal model, while those such as "the processes go parallel to each other and correspond to each other" suggest the non-causal representation. Even though Binet was quite successful with non-causal research he seemed to prefer causal reasoning. A causal understanding was the model Boring (1923, p. 35) intended when he remarked, "if the correlation is considerable, yet not perfect, say 60 percent, we say that the particular capacity is partly dependent on intelligence." "Dependent variable" is the terminology usually reserved for causal research. Intelligence test score is a non-causal variable. Boring's causal implication was, thus, most inappropriate. Incidentally, modern methodology would regard as equally inappropriate the view of a correlation coefficient as a percentage of variability accounted for.

Both Binet and Simon (1905a) as well as Boring (1923) emphasized the imperative interim status of the classification test score as a basis for understanding intelligence. Pearson (1930, pp. 56-57), on the other hand, stressed the clear logical difference between causal and non-causal models:

Formerly the quantitative scientist could only think in terms of causation, now he can think also in terms of correlation. This has not only enormously widened the field to which quantitative and therefore mathematical methods can be applied, but it has at the same time modified our philosophy of science and even of life itself.... The root idea at the bottom of correlation must not be treated as merely rebuilding on a securer mathematical basis statistical science. It is a much greater innovation which touches in its philosophical aspects the epistemology of all the sciences.

The tendency to ascribe causal characteristics to a non-causal variable is not unique to psychology (see Lave & Seskin, 1979). A more subtle misuse of a non-causal variable is a statement of the common-causes type suggested by Galton himself (1889, pp. 11; 196); not having identified any source of variability he was not logically justified in speaking of an empirical variable as a cause—common or otherwise. Messick (1975, p. 957) also may have intended the statement, "a measure estimates how much of something an individual displays or possesses" in the common-

cause sense. The nature of this language plus his added comment, "the basic question is, what is the nature of that something?" indicate the non-causal aspect of the measure: no identified, causal source of variability.

The Future

But why all the fuss? Surely everyone knows that correlation doesn't mean causation? The assessment of what everyone knows would be difficult indeed. Promoted perhaps by Galton's initial interpretation of common causes, many writers point out that a researcher should be wary in interpreting correlation coefficients. For example, the fact that there is a substantial correlation between the number of prostitutes and the number of churches in cities is sometimes used to illustrate how this coefficient may be misinterpreted because of a common variable to which both variables are related, namely, size of city. But in what sense is this a misinterpretation? A reader could get the impression that this caution means that there may be no relation between prevalence of prostitutes and preponderance of churches in the causal sense—an argument resembling an inverse of Galton's common causes. Correlation in no way empirically deals with identified causal variables, common or otherwise.

Whether common-causes are inferred from covariation or are used to argue against causal connections, Jensen (1969) would be fully justified on the basis of the "vast literature" (p. 81) in psychology in believing that many psychologists regarded IQ as an identified variable. Human rights, racism, lack of objectivity, are some public interpretations of what for many psychologists is simply an indiscriminate use of these two classes of concepts. The perceptive journalist, Walter Lippmann, expressed public concern regarding the way in which scientific views of intelligence were developing: "It leads one to suspect, after such a beginning, that the real promise and value of the investigation which Binet started is in danger of gross perversion by muddleheaded and prejudiced men" (Lippmann, 1922, p. 215). As was noted in the beginning of this paper (Peckham, 1979) and as Cronbach (1975b, p. 1) forewarned, the failure to recognize the structural difference in what we have designated causal and non-causal concepts has now led to the beginning of a legal obligation to desist. The public debate continues.

One step in the direction of improving public and scientific understanding, while still retaining the kudos that has accrued to psychology from the successful prediction of school achievement, is suggested in the title of a paper by Spearman (1931): "Our need for some science in place of the word 'intelligence'." Cronbach (1969) seemed in favor of at least getting rid of the name associated with the non-causal definition: "On the scientific side, it is vital to break away from such stereotyped terms as 'in-

telligence' and 'learning ability'" (p. 341). To Zigler's criticism of "unbridled environmentalists in whose writing the concept of capacity is treated as a dirty word," Cronbach (1969) replied, "But 'capacity' is a dirty word, incapable of being given meaning and overwhelmingly capable of confusing discussions. It and all words like it refer to nothing but an expectancy under present circumstances" (p. 346). If we were to detach the word intelligence from the non-causal operation, we would then be back to the Binet position of leaving it undefined and also of refusing to call the non-causal operation "intelligence." Unlike Binet with a label searching for an operation, we would have an operation searching for a label. We could use a neutral word such as *modulus*, indicating a number that depends on two or more other numbers, to identify all operations which have previously proved useful in predictions. For example, *B-1908* would be the modulus for Binet's operations that permitted some precision in predicting school achievement. Such a designation would not remove the logical limitations of a non-causal variable but would help to remove, for both scientist and public, some of the surplus meaning that was of concern to Boring (1923). Each child could then have a value of the *B-1908* modulus which could be used, "under present circumstances", to predict school achievement, if some precision in prediction and not a causal understanding was wanted. On the other hand, if the intention was to arrange a set of conditions to maximize level of school performance for a particular child (i.e., change the "present circumstances"), the *B-1908* modulus would not be needed. This option seems to have been Binet's preference (Binet, 1907, 1909, 1911; Binet & Simon, 1905a). Maximum level of school achievement would be determined by arranging optimal values of relevant causal variables.

Why do many scientific psychologists continue to argue with non-causal concepts as if they were causal? The logic is obviously faulty. Why are scientific psychologists persuaded by such arguments? Almost since the beginning of research in psychology there has been a failure clearly to distinguish between causal and non-causal models, a distinction which Binet sensed but could not articulate, which Galton articulated but somewhat misinterpreted, which Boring glossed but misrepresented, and which Pearson unambiguously differentiated. Shaver (1975, p. v) suggests that attribution processes are the "understanding of the causes of human behavior," motivated by a desire to understand and to predict. Perhaps persistent confusion of models might be understood on the basis of post hoc, ergo propter hoc (i.e., any observation permitting prediction is, ipso facto, a cause). Test scores that covary with school achievement are easily misinterpreted in the causal sense as measuring something. Thus, the typical research strategy has not been to regard covariation as encouragement to pursue causal studies in the common-causes sense, or

in any other sense, but rather to attempt to show that non-causal data can be interpreted almost as quasi-causal data.

Unlike Binet, we now have an abundance of multi-variable models but do little to pursue the numerous sources in causal research designs, including for example, obvious variables associated with the family (see Heslip, 1974). The erroneous use of non-causal concepts in causal arguments is unquestionably encouraged by a difference in the complexity of the concepts. Binet (1898, 1909) was interested in "individual psychology," by which he undoubtedly meant the identification of values of variables related to level of school achievement for *each* individual—a multi-variable causal problem. Having operationally identified the variables and empirically determined the functions relating them to school achievement, he would have sought to adjust the values of those variables susceptible to direct manipulation in order to maximize a particular student's level of school achievement. For example, heredity might be regarded as relatively immutable and teaching method as changeable within limits. A maximum level of school achievement for a particular pupil might be arranged by determining a teaching method appropriate for that pupil.

Individual psychology is to be distinguished from what has been labelled "the psychology of individual differences": a non-causal understanding which does not consider values of each of a number of identified sources of variability for one individual. Rather this non-causal understanding compares a number of individuals with regard to the correspondence *of* the resultant *of* the complexity *of* unknown values *of* unidentified sources of variability and their possible interrelations on a test variable with the resultant *of* the complexity *of* unknown values *of* unidentified sources of variability and their possible interrelations for school achievement: a non-causal problem appearing at least as complex as a causal one when expressed in causal terminology! However, when a non-causal concept is misused directly in the causal sense or in the common-causes sense, scientific misunderstanding of level of school achievement results.

Another possible explanation of the persistent confusion is suggested by Shaver's (1975) conceptualization of attribution processes together with the success of classification tests in schools and during World War I and World War II. The effective use of intelligence and pilot aptitude tests encouraged scientific and public acknowledgement of psychology as a science. Both models permit prediction but frequently a non-causal prediction was endowed with causal characteristics. With recognition largely for efforts with non-causal variables, scientific psychologists heeded but little the cautions of writers such as Binet and Cronbach; although it was generally stated that correlation did not necessarily mean

causation, non-causal concepts were treated as if they were causal—hence the “vast literature” to which Jensen refers.

Finally, a spirited and timely argument might be made for understanding the persistence of the confusion in terms of Cronbach's (1975b) suggestion that *zeitgeist* is a salient factor in the history of the development of mental tests. The criticism of Ross (1969) that *zeitgeist*, as used by Boring, was without an operational definition has been answered satisfactorily by providing such a definition (Hyman & Shephard, 1980): and conjectures such as those of Cronbach and that which follows here may now be examined. Hyman and Shephard's argument is that, at least with regard to classification test variables, the scholarly *zeitgeist* of psychology is irretrievably linked to the public *zeitgeist*. For example, as noted above, the confusion regarding models began with the search for a scientific definition of *intelligence*, a word lifted from the language of the general public (a common procedure in psychology). Intelligence was given the causal interpretation typical of non-scientific attribution processes and may continue in this fashion until changes in both public and scientific psychology *zeitgeists* experience a change in meaning of the word.

Perhaps the change will be in the direction of a scientific approach to understanding *each* child's behavior. Perhaps this reorientation will result in a reunion of mental tests with the mainstream of psychology as Anastasi (1967, p. 297) hoped. Hebb (1978, p. 1143) has said, “testing has clearly been on the side of the angels.” It was and still is, but the use of classification test variables as identified sources of variability is not—nor has it ever been.

References

- Anastasi, A. Psychology, psychologists, and psychological testing. *American Psychologist*, 1967, 22, 297-306.
- Binet, A. La mesure en psychologie individuelle. *Revue Philosophique*, 1898, 46, 113-123.
- Binet, A. Les nouvelles classes de perfectionnement. *Bulletin de la Société Libre pour l'Étude Psychologique de l'Enfant*, 1907, 41, 170-183.
- Binet, A. *Les idées modernes sur les enfants*. Paris: Flammarion, 1909 (1910 ed. used).
- Binet, A. Nouvelles recherches sur la mesure du niveau intellectuel chez les enfants d'école. *L'Année Psychologique*, 1911, 17, 145-201.
- Binet, A., & Henri, V. La psychologie individuelle. *L'Année Psychologique*, 1895, 2, 411-465.
- Binet, A., & Simon, T. Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 1905, 11, 191-244. (a)
- Binet, A., & Simon, T. Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence. *L'Année Psychologique*, 1905, 11, 163-190. (b)
- Binet, A., & Simon, T. Le développement de l'intelligence chez les enfants. *L'Année Psychologique*, 1908, 14, 1-94.
- Binet, A., & Simon, T. La mesure de l'intelligence chez les enfants (avec démonstrations). *Bulletin de la Société Clinique de Médecine Mentale*, 1909, 9, 1-10.

- Boring, E.G. Intelligence as the tests test it. *New Republic*, 1923, 35, 35-37.
- Boring, E.G. *A history of experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts, 1950.
- Bridgman, P.W. *The logic of modern physics*. New York: Macmillan, 1927.
- Commission des anormaux. *Bulletin de la Société Libre pour L'Étude Psychologique de l'Enfant*, 1904, 15, 406-408; 16, 429.
- Cronbach, L.J. The two disciplines of scientific psychology. *American Psychologist*, 1957, 12, 671-684.
- Cronbach, L.J. Heredity, environment and educational policy. *Harvard Educational Review*, 1969, 39, 338-347.
- Cronbach, L.J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, 30, 116-127. (a)
- Cronbach, L.J. Five decades of public controversy over mental testing. *American Psychologist*, 1975, 30, 1-14. (b)
- Cronbach, L.J., & Meehl, P.E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, 52, 281-302.
- Einstein, A. *The world as I see it* (A. Harris, trans.). London: John Lane, 1935.
- Galton, F. Family likeness in stature. *Royal Society of London, Proceedings*, 1886, 40, 42-63. Supplemented with an Appendix by J.D. Hamilton Dickson on Pages 63-72.
- Galton, F. Co-relations and their measurement, chiefly from anthropomorphic data. *Royal Society of London, Proceedings*, 1888, 45, 135-145.
- Galton, F. *Natural inheritance*. London: Macmillan, 1889.
- Galton, F. *Memories of my life*. London: Methuen, 1908.
- Goddard, H.H. Mentality tests: A symposium. *Journal of Educational Psychology*, 1916, 7, 231-233.
- Goddard, H.H. *Human efficiency and levels of intelligence*. Princeton: Princeton University Press, 1920.
- Harré, R., & Madden, E.H. *Causal powers: A theory of natural necessity*. Oxford: Blackwell, 1975.
- Hebb, D.O. Open letter: To a friend who thinks the IQ is a social evil. *American Psychologist*, 1978, 33, 1143-1144.
- Heslip, D.W. *Maternal response and arithmetic achievement levels of elementary school boys*. Unpublished master's thesis, University of Manitoba, 1974.
- Hume, D. *A treatise of human nature*. Oxford: Clarendon Press, 1960.
- Hyman, B.M., & Shephard, A.H. Zeitgeist: The development of an operational definition. *Journal of Mind and Behavior*, 1980, 1, 227-246.
- Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, 1921, 12, 123-147; 195-216; 271-275.
- Jensen, A.R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, 39, 1-123.
- Kendall, M.G. Studies in the history of probability and statistics: II. The beginnings of a probability calculus. *Biometrika*, 1956, 43, 1-14.
- Lagrange, J.L. Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations; dans lequel on examine les avantages de cette méthode par le calcul des probabilités; et où l'on résoud différens problèmes relatifs à cette matière. *Miscellanea Taurinensia (Mélanges de Philosophie et de Mathématique de la Société Royale de Turin)*, 1770-1773, 5, 167-232.
- Lave, L.B., & Seskin, E.P. Epidemiology, causality, and public policy. *American Scientist*, 1979, 67, 178-186.
- Lippmann, W. The mental age of Americans. *New Republic*, 1922, 32, 213-215.
- Messick, S. The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 1975, 30, 955-966.
- Michotte, A.E. *The perception of causality*. New York: Basic Books, 1963.
- Pastore, N. The Army intelligence tests and Walter Lippman. *Journal of the History of the Behavioral Sciences*, 1978, 14, 316-327.
- Pearson, K. *The life, letters and labours of Francis Galton* (3 vols.). London: Cambridge University Press, 1914-1930.

- Peckham, R. [Larry P. v. Riles.] Judge strikes down IQ testing. *APA Monitor*, 1979, 10, 8.
- Quetelet, L.A.J. *Sur l'homme et le développement de ses facultés, ou essai de physique sociale*. Paris: Bachelier, 1835.
- Ross, D. The "Zeitgeist" and American psychology. *Journal of the History of the Behavioral Sciences*, 1969, 5, 256-262.
- Shaver, K.G. *An introduction to attribution processes*. Cambridge, MA: Winthrop, 1975.
- Shephard, A.H. Toss of a coin—probability and games of chance: Mathematical models of physical phenomena. *Canadian Psychological Review*, 1978, 19, 105-115.
- Shultz, T.R., & Ravinsky, F.B. Similarity as a principle of causal inference. *Child Development*, 1977, 48, 1552-1558.
- Simpson, T. *Miscellaneous tracts on some curious, and very interesting subjects in mechanics, physical-astronomy, and speculative mathematics; wherein the precision of the equinox, the mutation of the earth's axis, and the motion of the moon in her orbit, are determined*. London: Nourse, 1757.
- Spearman, C. "General intelligence" objectively determined and measured. *American Journal of Psychology*, 1904, 15, 201-293. (a)
- Spearman, C. The proof and measurement of association between two things. *American Journal of Psychology*, 1904, 15, 72-101. (b)
- Spearman, C. Our need of some science in place of the word "intelligence." *Journal of Educational Psychology*, 1931, 22, 401-410.
- Terman, L.M. *The measurement of intelligence*. Boston: Houghton Mifflin, 1916.
- Tuddenham, R.D. The nature and measurement of intelligence. In L. Postman (Ed.), *Psychology in the making*. New York: Knopf, 1966.
- Varon, E.J. Alfred Binet's concept of intelligence. *Psychological Review*, 1936, 43, 32-58.
- Wolf, T.H. *Alfred Binet*. Chicago: University of Chicago Press, 1973.