

The pitfalls of "mental tests" / by Charles S. Myers.

Contributors

Myers, Charles S. 1873-1946.

Publication/Creation

[Place of publication not identified] : [publisher not identified], [1911?]

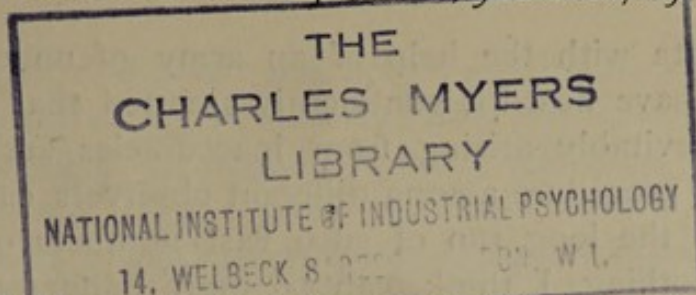
Persistent URL

<https://wellcomecollection.org/works/qrt8ct4q>



Wellcome Collection
183 Euston Road
London NW1 2BE UK
T +44 (0)20 7611 8722
E library@wellcomecollection.org
<https://wellcomecollection.org>

Reprinted from the *BRITISH MEDICAL JOURNAL*, Jan. 28th, 1911.



The Pitfalls of "Mental Tests."

BY CHARLES S. MYERS, M.D., ScD.

Lecturer on Experimental Psychology, University of Cambridge.

IN this country at least psychology seems likely to suffer from the same dangers of popularization as have for years past been affecting anthropology. Just as it has been widely supposed that only a printed book of instructions and queries need be read for an amateur to sally forth into the field and collect reliable physical measurements or trustworthy evidence of social organization, so there appears to be starting a belief that no special course of training is necessary in order to conduct on a large scale investigations of a psychological nature.

Folk are loth to recognize that the younger sciences—for example, economics, genetics, psychology, education—demand as much adequate preparation and study as the older before reliable work can be undertaken in their respective fields. Certainly every one think himself capable of discoursing and deciding about themes of psychological interest. Royal and Departmental Commissions, which do not contain among their members a single psychologist are appointed to report on matters which are fundamentally of psychological concern. The psychologically untrained physician does not hesitate to pronounce on the psychology of insanity, nor the biologist on human and animal intelligence and instinct.

I want to protest as strongly as I can against the notion that any useful purpose can be served, so far as psychology is concerned, by collecting masses of psychological

data with the help of an army of untrained observers. I have heard it confidently asserted that the gross errors inevitably arising from inaccuracies and inconsistencies of procedure among different observers cancel one another in the long run of such vast numbers of measurements. Nothing, I think, can be more dangerous or false than this idea that the untrustworthiness of crude methods diminishes as the number of observers increases. It involves the assumption that in the long run errors occur to an equal extent in opposite directions—a most unlikely hypothesis.

Individual differences in mode of measurement are great enough even in anthropometry, despite the standardization of measurements. There is, I believe, a well-founded rumour that when the pigmies visited this country and were independently measured by several observers well practised in anthropometry, the results obtained by these observers were startlingly divergent. If this be so, the sooner the fact is put on record the better for the future security of anthropometry. In any case I am sure that hitherto it has not been adequately realized how untrustworthy is a comparison between the measurements obtained by different observers upon the same individual. I can only give one or two actual examples, but these, perhaps, are sufficiently striking. Professor Cunningham and Mr. Browne measured the heads of several anatomists at a meeting of the Anatomical Society some years ago. One of their subjects gave a head length of 198 mm. and a head breadth of 147 mm.—that is, a cephalic index of 74.24. The same individual happened to be measured subsequently by Miss Alice Lee, fully as competent an observer, who obtained a head length of 195 mm. and a head breadth of 150 mm.—that is, a cephalic index of 76.92. Thus there was a divergence of over 3.5 per cent. in the values of the index obtained from the same individual by these observers, although each claimed to be taking precisely the same measurements—the maximal head length and the maximal head breadth.

A similar experience befell me only a short time ago. My head was measured by an observer who had for some months been engaged in making a vast collection of anthropometric data. He entered my head length as 202 mm., my head breadth as 168 mm.—that is to say, a cephalic index of 83.2. Doubting whether these data were accurate, I took the calipers in my own hands and obtained a head length increased by 4 mm. and a head breadth diminished by 2 mm., yielding a cephalic index of 80.6. The latter I know to be approximately correct.

If inaccuracies to this extent occur when anthropometry is in the hands of fairly-trained observers, what must be their size when the measurements are undertaken by the interested amateur! And, if they are great in measuring the *physical* characters of man, what must they be in measuring the *mental* characters! For here we have not only the dangers arising from the improper use and reading of the instrument, but also the different effects upon the subjects' mental condition produced by different observers. One observer knows his subjects well, another awes them by an unsympathetic attitude, while another unconsciously helps them by suggestion.

We have the further difficulty that it is impossible as yet to standardize mental tests as we have standardized physical measurements. Far more laboratory work is necessary before such fixity becomes possible or desirable. The approved test of to-day is rejected to-morrow.

But I will leave these difficulties on one side and pass on to the *purposes* of this wholesale application of "mental tests." One of the objects is to discover by statistical means the differences which exist in different communities. A vast number of measurements of a given character is taken in one community and an equal number of measurements of the same character is taken in another. The averages of the two series are compared, and the conclusion is drawn that an undoubted, or a probable, or no certain difference exists between the two communities in respect of this character. As is well known, the certainty

of the difference depends not only on the number of observations but also on the relation between the amount of the difference and the uniformity of the individual measurements within either community. If the individual measurements within a community differ widely from one another, it is obvious that the difference between the averages must be proportionately wide, in order to be certain that it is not merely accidental. Now, so far as the physical measurements of mankind go, they do differ enormously within a given community. They differ so widely that it may be said that statistics can seldom give us any *new* information as regards racial differences of measurements. Statistics serve, as a rule, but to establish and to give a measure of *observable* differences. Statistics cannot inform you that one community has broader noses, darker skins, more curly hair, or greater stature than another, unless that character is manifest enough to be apparent to the non-statistical eye. And if this be true as regards physical characters, it holds yet more strongly in respect of mental characters, inasmuch as they exhibit still wider individual variation within a community. For these reasons we must be chary against expecting from statistical manipulations more striking results than from the very nature of the data they are capable of yielding.

I am fully aware that these opinions savour of heterodoxy. I shall be asked, Are those not striking and new results which have been lately reached by statistical methods, showing the absence of correlation between the state of nutrition of school children and their mental capacity, or between alcoholism in the parent and defective health in the offspring? And I reply that, in my opinion, these results have no real value. They have been obtained by applying scientific methods to the solution of a problem of such complexity that the solution appears in the form of a meaningless blur.

The plain man believes that "one can prove anything by statistics," and I fear that such time-worn sayings have a certain basis of truth. For the wholesale collectors

of measurements are apt to pay too little regard to the complexity of the conditions influencing the problem and the material which they are gathering. All they desire is an enormous mass of data, and these—good, bad, comparable, and non-comparable—they pour into the statistical mill with the object and result of arriving at conclusions statistically invulnerable. Into this mill, for instance, they pour all the data concerning the relative efficiency, physical and mental, of the children of drunken and sober parents, practically regardless of the question as to whether the parents are strong, robust folk who are abstemious, or, say every Saturday night, regularly indulge in intoxication, or whether they are feeble workmen of the sober “good young man” type, or weakly degenerates, inheriting and transmitting disorders of nervous instability so closely associated with tendencies to crime and to chronic alcoholism.

Similarly, in dealing with the relation of mental capacity to bodily nutrition among children, a recent investigator has been content to take the verdict of school teachers on the mental capacity of their school children, with the result that some teachers classified 33 per cent., others only 1 per cent., of the children as brilliant! On the ground that the ablest parents can provide the best nutrition and transmit their ability to their children, it might be argued *a priori* that dull children would not be so well nourished as bright children. And this conjecture is supported by earlier trustworthy evidence. On the other hand, its *raison d'être* can only hold for the lower and lower middle classes, and we have the further complication that unwholesome food may produce a false appearance of good nutrition. Apart from food supply, exceptionally bright children are worse nourished, and exceptionally dull children are better nourished, than the average child. These complications should no doubt be taken into consideration, as should such factors as employment outside school hours and the mental and physical condition of the parents. When all these counter-

acting influences are thrown into the statistical melting pot, is it surprising that the result is a mere blur, showing an absence of significant correlation, a small correlation in one direction in one school and in an inverse direction in another?

This neglect to analyse and to take heed of what is actually being measured is specially prone to occur in the use of mental tests. In other sciences there is little or no real difficulty in observing what we are measuring, if only the experimenter take reasonable care. But in psychology we can only ascertain what we are testing by recourse to introspection on the part of the subject. To neglect introspection in psychological experiment is usually to court certain disaster. If we are in total ignorance of what has been going on in the mind of the subject during the experiment, it is rarely possible to argue from the objective data—from the measurements which it yields. For example, we may be trying to determine whether any correlation exists between sensory discrimination and general intelligence. A positive result may be simply due to the fact that the very nature of the test has compelled the subject to use his intelligence while carrying out sensory discriminations. We may be correlating mental ability with mental fatigue, and neglect the fact that sometimes we may not be measuring fatigue at all, that in some subjects the task becomes automatic, in others tedious, or that boredom may be in others overcome by motives of duty or ambition. We may be testing the visual acuity of two persons and obtain a different result from each, despite the fact that really they have the same visual acuity. The result may be due to the fact that the one subject strains every effort to interpret what he but dimly sees, while the other only reads what he believes he can clearly see. Thus again we merely obtain a blurred or erroneous result from the blind applications of statistical methods to measurements which are really meaningless owing to our failure to analyze the conditions determining the character we are measuring.

The danger of drawing conclusions from too small a number of subjects is well illustrated in the results of an inquiry recently conducted into the correlation between ability in mathematics and ability in classics in the various forms of a public school. In the highest form the correlation was found to be $+0.20$; but in the following year in the same form it amounted to $+0.52$. In the form below it was $+0.23$, in the next lower form $+0.76$, and in the next -0.25 .

But there is likewise a pitfall from the use of a large number of subjects, and this I will illustrate, as before, by analogy from physical anthropology. It is obvious that if you determine the correlation between head length and head breadth for one race or for one ethnic element of a mixed people—for example, the Cornishman in our own country—we shall find this correlation to be quite different from that obtained from another race or for another ethnic element of a mixed people—for example the East Anglian. There can be no question about the existence of similar difficulties with regard to correlation of mental characters. How wide the racial differences are in the correlation of mental characters is, of course, unknown. But no doubt at one time in a given class or school of our heterogeneous population the ethnic diversity may be small, at another time it may be great. At one time one racial element may preponderate, at another time another. This possibly is one explanation of the marked discrepancies obtained by different observers and by the same observer at different times, using the same mental tests as far as possible in the same manner. It provides yet another confirmation of my thesis that the wholesale collection of measurement is apt to give us only a very blurred and often a very inaccurate picture of the factors which really underlie the problem under investigation. To sum up, it does not give results of psychological value, because the psychological standpoint, the experience of the individual, is neglected. It is only too apt to obscure actual correlations or to reveal spurious correlation because

insufficient care is taken to analyse the conditions which are really at work during the experiments. It leads to inaccurate results owing to the errors arising from individual differences in applying the tests.

For these reasons I urge extreme caution, at least for the present, in standardizing "mental tests" and in popularizing their use. In some forms, no doubt, tests can be usefully applied *en masse*—for example, with the object of determining the standard of intellect which a boy of given age should attain in order to class him as suitable or unsuitable to be taught in an "ordinary" or a "special" school. But such tests are "tests of production," not "mental tests." They determine how *much* an individual can work, how *much* he knows—not *how* he works, *how* he knows. A man's productivity, of course, is what we want to ascertain in everyday life. We do not care how a man comes to use or to acquire his powers; we are content with a mere dynamometric or other record of his prowess. From this aspect, mass experiments must have *some* value. But this aspect cannot properly be called the psychological aspect.