

That is, if we took the job of a writer, or took the job of a Congressman, someone has to sit down and first decide what makes a good writer and what makes a good Congressman. If you can't decide on that it is of absolutely no value to know how sociable a man is, how direct he is, or how indirect he is. Unless you know for sure that you want decisiveness and in what dimensions you want it, there is no sense in testing anyone. In psychological language this is called criteria.

We know nothing about the criteria of any job. This is indicated by the fact that when the same corporations that use this type of testing thought they might as well find out if they knew what they were doing. They asked the Educational Testing Service to do experiments on criteria of executives. This included 14 major firms such as IBM. The Educational Testing Service set up a reasonable experiment. A man working as an executive, with an in-box and out-box, did everything he was told to do. The conclusion that the Educational Testing Service fed back to the corporations is that they did not know anything, nor could they find out anything, about the criteria of executives. Executives vary in their personality and behavior.

People who observe notice this. But testers would like, of course, to be more specific because this enables them to make profitable claims.

In testing, it is important to understand technical phrases such as reliability and validity. Reliability is a psychological phrase that can best be explained by using the thermometer as an example. If a person is without fever, and his temperature is taken with a good thermometer, by a good physician, the thermometer registers approximately 98.6.

If the thermometer in use was unreliable, and the person had no fever, yet it registered anywhere between 90 and 110, then the physician could not use this thermometer for evaluation. Measuring just one or two points of temperature is essential. If your temperature is 97, it could be quite dangerous, and 101 could be the beginning of a heavy infection. If the thermometer was inaccurate in its reliability, then it could not be used as a measuring instrument. No validity could result from it because of its bad reliability.

Reliability is not the accuracy of the prediction, but of the measuring instrument, such as a ruler. If it is ostensibly 6 inches long and is really only five and a half inches, and you measure your height with it, you come out badly.

The manuals generally claim high reliability for tests, whether it is the Washburne or Edwards or MMPI. This is done by the test authors, who have made up the test, and have a natural human tendency to come out with the best possible results.

Later tests on test reliability, called test-retest reliability, are less optimistic. This is a most stringent method that gives the same person the same test over a period of time, perhaps a few weeks apart. This method indicates that most of these tests are not reliable; that is, the same person, 2 weeks apart, scores quite differently on the same test.

His mental health hasn't changed, his sociability hasn't changed, nothing particular about this man has changed in a few weeks. What has changed is the test score, which means that the tests are not reliable. As an example, the Thurstone temperament schedule, which is used