Summary of Test Development

By: Kimberly R. Brinkmeyer, Ph.D.

Personality assessment is useful for describing an individual on characteristics which can not be directly observed. Behaviors are visible to people, but the reasons behind them and the motivations for them are not observable. Psychological assessment results provide a vocabulary for describing propensities and a view of the "whys" behind the behaviors. This information sets the stage for more effective employee and manager selection, succession planning, team building, and professional development.

During test development, the first question to answer is "what to measure". Items are then written to reflect the behaviors associated with the dimension(s) being measured. For example, extroversion is defined as: interest in or behavior directed toward others. Thus, these items might be written to reflect the degree to which a person enjoys (or does not enjoy) social settings, crowded events, and making presentations. Sample items focusing on extroversion might be "in a group, I enjoy attracting attention to myself" or "I don't care for large, noisy crowds."

Once a pool of test items are written—pool size is determined by the number of dimensions being assessed and the number of items deemed necessary to tap into each dimension—they are pilot tested to determine the degree to which they correlate with each other, they differentiate people (some endorse, some do not), and they are reliable. Reliability is determined by test-retest—subjects answer the question the same way on multiple administrations—and by internal consistency—items which are designed to measure extroversion tend to correlate, or hang-together.

The next step is to determine the validity of the measure, which is a bit more complicated to explain. First, there is what's known as the test-test validation process which correlates (see attachment for explanation of correlations) scores on our instruments with other instruments. These test-to-test correlations are conducted with instruments that are hypothesized to have similar or related constructs and with instruments that are hypothesized to be unrelated. For example, the process of validating the Character Assessment included having subjects take the Character along with the ASVAB, PSI Basic Skills Test (both should be unrelated), Myers-Briggs, SDS, Interpersonal Adjective Scales, Big Five Factor Markers, and the MMPI-2 (all of which should have some relationship to the measures). These analyses resulted in correlations that confirmed hypothesized relationships.

The next level of validation should include correlations between test scores and relevant non-test indicators—such as actual performance ratings. This step is taken to validate (confirm or not) whether the instrument accurately measures the predicted behavior and the impact on performance. For example, those who have a high scores on the CDR Character Assessment "Adjustment" scale and a high CDR Risk Assessment "Egotist" scale will generally have higher self-ratings on 360 performance reviews. This translates to people who have higher opinions about their own performance in comparison with the perceptions of others. Thus, the correlations will be higher between these scale scores and the resulting behavior ratings. The validation process is not simple and it is important to perform statistical analyses using a variety of non-test indicators and performance results. In addition to performance reviews, other examples of non-test indicators may include: sales results, customer retention, customer complaints, accidents, turnover, errors, etc. We can provide summaries of this analysis or actual sample validation studies conducted for clients.

When evaluating personality assessment measures or styles inventories, it is important to determine whether the assessment authors perform only the first level of validity analysis, i.e. test to test, or, also validated the assessment results through correlations with actual performance behaviors. The test development process determines the applicability of the assessment results to workplace decisions. As with our assessments, only valid and reliable tools, as determined through the test development process, are valid for selection decisions. In other words, our measures correlate to actual results.

What is a Correlation?

Correlation is a measure of how closely two variables move together through time, or the degree to which two variables are associated. A positive correlation exists when two variables increase or decrease together. For example, height and weight are positively correlated, meaning that as height increases, so does weight. More of one means more of the other. A negative correlation exists when increases in one variable are accompanied by decreases in the other, and vice versa. For example, research might show that self-esteem and depression are negatively correlated, indicating that as self-esteem increases incidence of depression decrease. More of one means less of the other.

What Does a Correlation Represent?

There is a simple technique for illustrating the real size and importance of correlations. The "binomial effect size display" (BESD[*]) allows correlation coefficients to be interpreted in terms of the percentage of correct classifications they represent.

Effect (Impact) Sizes Corresponding to Various Values of Correlations

Correlation	Success Rate Increased	Difference in Success Rates
.10	From .45 to .55	10%
.20	From .40 to .60	20%
.30	From .35 to .65	30%
.40	From .30 to .70	40%
.50	From .25 to .75	50%

The question addressed by BESD is: what is the effect on the success rate (i.e., selection rate, improvement rate) attributable to a certain action or variable? So, when the correlation coefficient is the result of sound research, the increase in success rate of a procedure will be roughly equivalent to the value of the correlation coefficient.

[*] Adapted from "A Simple General Purpose Display of Magnitude of Experimental Effect, by R. Rosenthal & D. B. Rubin. *In Journal of Educational Psychology*, 74, 166-169.