Predictive Index® Research Programs

A Report on Reliability and Construct Validity

SUMMARY

J. Christopher Perry, M.P.H., M.D.
Philip W. Lavori, Ph.D.
September 25, 1983

PI Worldwide

SUMMARY

In January, 1982 Praendex Incorporated, parent company of PI Worldwide, commissioned an independent professional study of the reliability and validity of the Predictive Index® (PI), a free-choice inventory for the objective measurement of a broad range of work related behaviors and potentials based on measurement of four primary personality traits. Scored on three scales (Self, Self-Concept, Synthesis) the Predictive Index distinguishes between an individual's basic personality traits, adaptations to his or her environment, and actual behavior (performance) in the work environment.

Drs. J. Christopher Perry and Philip Lavori of the staff of the Harvard Medical School agreed to undertake and collaborate on the project with the understanding that they were to have complete, independent control of its design and execution, with their findings to be presented in a report written exclusively by them. Praendex's only role in the project was the gathering of raw data requested by the authors.

The project was carried out over the next 18 months with Drs. Perry and Lavori reporting the results of their statistical analyses and their conclusions in September, 1983. (A copy of their complete report is available on request.)

The defined objectives of the study were threefold:

1. To determine the reliability of the Predictive Index's measurements of personality traits

2. To determine the accuracy of the norms and scores used to quantify those measurements

3. To determine the extent to which the Predictive Index's measurements of personality traits correlate with another standard personality inventory, in this case the measurements of personality traits made by R. B. Cattell's "16 PF Personality Factor Questionnaire".

What follows is a brief summary and explanation of the results of the study; quotations in this summary are taken directly from the report.

RELIABILITY

Internal Consistency

To establish the usefulness of a test which measures personality traits, two measurements of reliability should be studied. The first is a measurement of internal consistency. In the case of Predictive Index it was necessary to determine if all of the adjectives in the checklist classified as measuring a particular trait (either Factor A, B, C or D) all do in fact measure that trait.

In that regard, the report states that *"all of the scales have reliability coefficients above .70 - the acceptable cutoff - while 7 of the 8 scales [Self A, B, C, D and Self-Concept A, B, D] have reliability coefficients above .80 and are thus highly acceptable."*

Stability

The second reliability measurement is stability, or the degree to which scores change significantly on the re-test of an individual. If there are significant differences between the two test scores, particularly when test and re-test are administered within a short period of time, it is impossible to determine which test is the more accurate.

It was expected that the re-test stability figures would be highest for the Self pattern, since that scale taps personality traits directly. The Self-Concept pattern measures an individual's response to environmental influences, and since environments, and perceptions of them, change, lower correlations were expected in re-tests. The Synthesis pattern, a composition of Self and Self-Concept, was expected to yield reliability figures somewhere between the levels of the figures for Self and Self-Concept.

In the case of this study, re-tests were administered between 3 months and 8 years of initial administration. Re-tests between 3 months and 2 years had the highest re-test reliability. The average coefficient for the four Self factors was .65, for the four Self-Concept factors .63. After two years the consistency of re-test results slowly began to decline and those re-tested after 4 to 8 years show an overall average reliability coefficient of .50, which the study found *"quite remarkable given the length of time of the re-test interval."*

The study also analyzed short-term (less than 3 months) re-test reliability and found: *"As would be anticipated, the estimated short-term test-re-test reliabilities are generally higher for all four Factors and for the Norm M. These are in the same range as the short-term stability coefficients of the 16PF."*

ACCURACY OF SCORES

Means and Standard Deviations

In the study average scores were computed based on the test responses of 260 people employed in the commercial/industrial sectors of the United States and Canada. The scales on the three patterns approximate a normal distribution, although for both the Self and Self-Concept patterns, the A and B scales are slightly skewed to the high side, while the C and D scales are normally distributed.

The report states that *"The plotted means on the P.I. Data Sheet [scoring form] are very close to demonstrating a vertical profile and are well within the sampling variations. The size of these standard deviations is due to variability of the job types. This further emphasizes the value of specific job studies."*

Gender Differences

The study found that in all three patterns men generally exhibited higher A-Factors than women, while women generally exhibited higher C-Factors in the Self-Concept and Synthesis. However, when survey results of men and women were adjusted for the job they held these differences diminished. *"This finding probably reflects the job differences even within the same job category level [e.g: car salesman versus secretary]. It does dramatically point out that the most valid comparison of P.I. scales is between an individual and the performance norm for the exact type of job that he or she will be doing. Sex differences are much less important."*

INDEPENDENCE OF PREDICTIVE INDEX FACTORS

In order to determine the degree to which the four Primary Factors of the Predictive Index are independent of each other, intercorrelations among all four Factors were done. The variances found are reported in detail in the complete report of the study.

The analysis found that fairly high interrelations exist between the A and B Factors on one hand, and between the C and

D Factors on the other. Those relationships are to be expected since the A and B Factors, which both measure "proactive" drives, tap some personality traits common to both, while the C and D Factors, both of which measure "reactive" drives, tap some traits common to both of those Factors.

On the subject of those interrelationships, the report of the study states, *"In this case it is not a weakness of the instrument, but simply a reflection of the fact that many aspects of ordinary and describable personality are not 'independent', but are overlapping and correlated...One is certainly better off with a few well chosen, overlapping variables with simple, intuitive interpretations than with hybridized constructs whose sole claim is that they satisfy the psychometric ideal of independence."*

To further explain the terms used above, proactive behavior is characterized by action upon the environment, changing it by impacting or influencing it, thus producing change in either the condition of things and the way they are done, or in the attitudes or behavior of people. Reactive behavior is characterized by acceptance of the environment as it is perceived, and accommodation to it, effecting no change in its condition.

Among all four Factors there are major differences which make pertinent distinctions in work behavior. For example, the A and B Factors distinguish between dominant (authoritative) versus persuasive (delegative) management styles, while the C and D Factors distinguish between passive versus conscientious work behavior, Many other examples of such distinctions are found in the complete report and in the manuals of the Predictive Index and the 16PF.

CONSTRUCT VALIDITY

While reliability refers to the consistency of a measure, validity refers to the accuracy of a measure. A measure is valid if it actually measures what it purports to measure. The validity of a measure can be assessed in a number of ways. Construct validity is demonstrated when a measure is statistically compared with another measure of similar and/or different concepts. To be successful, the comparison measure must have been soundly constructed and be generally accepted. Such research has been successfully performed twice, in the Perry and Lavori study and in a more recent study called "The Construct Validity of Predictive Index and 16PF". Both of these studies compared the PI with Cattell's 16PF (The 16PF is a well-respected and well-researched personality assessment). A construct validity study involves looking at patterns of correlations. Correlations are mathematical measures that can identify the presence and strength of the relationship between two variables.   A pattern should emerge that meets the following expectations: Factors that are defined in a similar way by both the PI and the 16PF should prove to be very mathematically related (e.g., PI's extroversion factor and the 16PF extroversion factor), and factors that are defined in a dissimilar way on both the PI and the 16PF should prove to be mathematically unrelated (e.g., PI's extroversion factor and the 16PF emotional stability factor).

In both of these studies, the PI very successfully demonstrated construct validity: the relationships you would intuitively expect should be related were mathematically related, and the relationships you would intuitively not expect should be related were mathematically unrelated.