



**PARADIGM**  
PERSONALITY LABS

**RELIABILITY AND VALIDITY OF  
THE WORKPLACE BIG FIVE PROFILE™**





## RELIABILITY AND VALIDITY OF THE WORKPLACE BIG FIVE PROFILE™

Today's organizations and leaders face a demanding challenge in choosing from among thousands of personality assessment products and services. Personality testing is a \$500 million industry with an estimated annual growth rate of 10%. With more than 2,500 personality questionnaires on the market today and dozens of new companies appearing annually, the challenge of finding the right assessment grows more difficult every day.

Validity and reliability are critical factors for our clients. Why? Because only valid and reliable assessments provide proven, accurate insights that advance business goals, making them a worthy investment. Paradigm Personality Labs' assessments are cost-effective, efficient, and legally sound. Moreover, the WorkPlace Big Five Profile™ is one of the most psychometrically robust tools on the market, maintaining an average coefficient alpha of .83. It also complies with all International Test Commission guidelines.

Our clients want to know how well the instrument measures what it proposes to measure and how well it predicts an organization's desired outcomes. This paper is for the informed decision-maker and contains information that expert psychometricians typically use when evaluating assessments.

### Current Standards

We use the most current standards to ensure that the WorkPlace Big Five Profile™ is valid and reliable. Accordingly, we define validity as "the degree to which evidence supports intended uses of a test" (Standards for Educational and Psychological Testing, 1999 & 2014). Our methods are also aligned with the Society for Industrial and Organizational Psychology which updated its Principles for the Validation and Use of Personnel Selection Procedures (2003) to be consistent with these established standards.



## What Does it Mean to Call a Test Reliable?

Reliable tests get consistent results when the tests are repeated. Reliability can decline with time, however, because experience, context, and maturational factors can influence participant responses to behavioral questions.

Test/retest studies are one common approach for assessing reliability.

- **Short-term test/retest:** In a short-term test/retest, the assessment is first administered at one point in time and then administered again to the same individuals one to three months later. Well-constructed tests should yield short-term test/retest reliabilities approaching .90. Previous internal research yielded an average short-term test/retest reliability of .88 for the WorkPlace supertraits. Updated test/retest research will be completed following the release of our latest assessment.
- **Long-term test/retest:** In a long-term test/retest study, the assessment is first administered at one point in time and then administered again to the same individuals one to three years later. Well-constructed tests should yield long-term test/retest reliabilities of .70 or higher. Previous internal research yielded an average long-term test/retest reliability of .72 for the WorkPlace supertraits. Updated test/retest research will be completed following the release of our latest assessment.
- Perhaps the most common measure of reliability is the coefficient alpha (often referred to as Cronbach's alpha). The coefficient alpha measures the degree to which a set of scale or test items are answered consistently. Alpha values of .70 are generally considered acceptable and satisfactory; alphas below .50 are usually not acceptable. Alpha values above .80 are regarded as quite strong. Alpha values above .90 reflect exceptional internal consistency although scores this high can also indicate that the scale items measuring the construct are too similar. The mean WorkPlace Big Five Profile™ subtrait coefficient alpha is .72 and the mean supertrait coefficient alpha of .83, both in line with established industry performance standards. Moreover, the application of more recent statistical innovations in reliability measurement places the subtrait and supertrait mean reliabilities at .77 and .88, respectively.



## What Does it Mean to Call a Test Valid?

Validity is a key issue for any academic reviewer of a personality questionnaire. It is defined as "the degree to which evidence and theory support the interpretation of test scores for the proposed use of tests" (Standards, 2014, p. 11).

From the user's perspective, validity is measured simply by the assessment's ability to accurately predict what it claims it will predict. Validity is an accumulation of evidence, and most organizations expect assessments to have published validity data. Instead of informing solely on content, construct, and criterion-related validity, we use modern psychometric standards that include intended purpose and business context in validity discussions.

We took the following issues into consideration in assessing the validity of the WorkPlace personality assessment:

- **Assessing for disorders:** The courts have ruled that assessments whose results provide information about psychological disorders are discriminatory and therefore cannot be used to make hiring decisions. The Americans with Disabilities Act considers such diagnostic assessments to be medical examinations, permissible only after the individual has been offered a job. The courts have said tests that measure normal personality are permissible if validity studies have proven the tests are relevant to the job. The WorkPlace neither measures for disorders nor reports them.
- **Bandwidth:** In testing, bandwidth refers to the scope of a specific measure. For instance, "Overall IQ" has broad bandwidth while "3-dimensional spatial rotation ability" has narrow bandwidth. In the context of WorkPlace, the Extraversion supertrait, for example, has broader bandwidth, while the subtrait E2 (Sociability) has a narrower bandwidth. Although the broader strokes of WorkPlace supertraits can help new users orient to personality constructs in general, specific roles and behaviors at work are typically best explained and understood by the more targeted bandwidth of subtraits. Efforts to understand performance and fit for sales for example are made more effective and actionable by focusing on multiple specific subtraits



such as low N4 (Rebound time) and high E2 (Sociability) compared to the general supertrait of Need for Stability and Consolidation. For this reason, we typically focus our discussions on WorkPlace's 23 subtraits rather than the five broad supertraits.

- **Compliance with employment law:** The original item list for the WorkPlace was more than 800 statements. In consultation with our labor/employment attorney, we have eliminated what could not be asked of a prospective employee before hiring. The WorkPlace has no items inquiring about political, religious, or social beliefs. The WorkPlace assessment has not been challenged, and to our knowledge, no Big Five test has ever "gone to court."
- **Item content:** The language of test items should reflect the context in which results will be applied. For the WorkPlace, this applies in two ways:
  - *Work context:* The WorkPlace language for test items is workplace language and the results are interpreted in terms of people's behavior at work. We use language such as "work in solitude" and "imagines new business concepts".
  - *Global/cultural context:* WorkPlace has been translated into several other languages. In each case, we translate into the target language, then have a different linguist translate the items back into English. We then "jury" that final translation to ensure the content matches the construct each item measures. We resolve discrepancies by creating new items in the target language that are faithful to the construct. Hence, each translation uses language that is natural for the target culture(s) to reflect the construct being measured.
- **Item format:** Studies have shown that items worded in the third person without pronouns (as in "Is a talker" or "Interrupts others") elicit a wider range of responses than items beginning with the personal pronoun (as in "She/he is a talker" or "I am a talker"). For this reason, we use the third person without pronouns throughout the WorkPlace.



- **Normative vs. ipsative.** Ipsative scales ask individuals to make a forced choice between a set of options. For example, an ipsative measure might ask “Which of the following best describes you?” followed by a pair of options such as “Driven” v. “Unambitious”. While ipsative measures may be useful, they only tell us about the relative preferences within an individual. As a result, ipsative scales are purely self-referencing and cannot help us make comparisons between people. To help us understand individual differences, we instead need a normative scale. Normative assessments gather responses from multiple items on a single scale and then analyze the results to determine how that individual's score on that scale compares with scores from others in the sample of interest. The WorkPlace is therefore a normative assessment specifically tailored to the work environment. Most ability, aptitude, and personality assessments are also normative assessments.
- **Norms.** Tests should be normed on the same kind of population that will use its results. To help create a well-balanced norm group, we modeled the distribution of our sample on data from the U.S. Bureau of Labor Statistics for employed individuals between the ages of 20 and 64. The characteristics of the final U.S. norm group sample of 2,840 individuals (Female = 1,331, Male = 1,509, mean age = 41.2) therefore match the BLS distributions and provide a principled path to maximizing the generalizability of our assessment to both current and potential future employees. This sample also covered 20 major industry segments, including finance, health care, retail, and manufacturing. Additional details on the norm group and related statistical considerations are provided in the WorkPlace Big Five Profile™ Technical Manual.
- **Predictive power.** Across all studies we have done with the WorkPlace, we have found that individual traits typically correlate from .15 to .30 with performance criteria. Predictive models that include multiple traits (e.g., using the subtraits N4 (Rebound time), E1 (Warmth), E2 (Sociability), E3 (Energy mode), A3 (Humility), and C5 (Methodicalness) to predict the criterion (e.g., sales volume) can achieve stronger relationships. As traits are only a part of the total person, we recommend combining trait scores with other predictors relevant to the performance



criterion being measured. Other such predictors may include job performance simulations and other competency-based assessments, experience factors, and relevant background checks (e.g., credit, police, academic). In most circumstances, situation-specific evidence is most informative. Please note that such studies are conducted at client organizations to include the job, industry, and business context. We do not share their confidential or proprietary information.

- **Response options.** Studies have shown that the use of all positive anchors (as in 1 through 5) with a Likert-type scale (e.g., strongly disagree through strongly agree) fails to elicit as wide a range of responses as do a mix of negative and positive anchors (as in -2 through +2). For this reason, we use the -2 through +2 format in the WorkPlace.
- **Results match predictions.** Across a variety of industries (e.g., banking, entertainment, government, manufacturing, utility, transportation), using WorkPlace results together with other traditional criteria to select employees leads to the predicted, effective workforce results including:
  - Reduced employee turnover
  - Increased workplace performance
  - Improved information for the acquiring manager to use in coaching and team building.
- **Social desirability.** Many tests have validity scales or social desirability scales (i.e., "lie scales" to determine whether the respondent is being truthful. Research indicates that such scales do not work. We follow the suggestion of Costa and McCrae (1992), who assert that the use of raters can control for socially desirable responses in the case of high-risk assessments and that careful attention to instructions given to respondents can minimize socially desirable responses. In rare circumstances, however, the tendency to self-enhance is so widespread that such additional steps make little difference. The WorkPlace questionnaire therefore also



includes an honesty pledge in which respondents must agree to truthful responses before being allowed to proceed with the assessment. In addition, the WorkPlace Consultant's Report provides a flag for any of six distinct "response sets" (or unlikely response patterns) that suggest the possibility that the respondent is not answering truthfully and could invalidate the results. For example, the response set "Tendency to Agree" is flagged whenever the respondent agrees or strongly agrees with 65 or more of the 93 WorkPlace trait items; less than 1% of the norm group agree to that extent.

## What Do You Want to Measure?

Identifying the proposed use of the assessment is critical to establishing predictive validity. Organizations have successfully used assessments based on the Five-Factor Model (including the WorkPlace) for various business purposes including:

- Assessing person-job and person-culture fit
- Coaching and career development
- Diversity training
- Employee engagement
- Hiring and selection
- Leadership development
- Performance development
- Personnel selection
- Project team design
- Research and validity studies on job fit
- Succession planning
- Team building



## Why Choose WorkPlace Big Five Profile™?

Our clients choose the WorkPlace Big Five Profile™ as their assessment for many reasons including:

- Based purely on the Five-Factor Model, the preeminent model for measuring, interpreting, and understanding personality. Over 30 years of applied and academic research has shown that the Five-Factor Model is universal and robust across industries, observers, geographic locations, languages, and gender.
- Maintains strong alpha coefficients at both the subtrait and supertrait levels.
- Provides accurate, accessible insights into complex human behavior without oversimplifying.
- Adheres to guidelines from both the International Test Commission and the Standards for Educational and Psychological Testing.
- Embedded throughout our extensive array of products, special reports, and forward-thinking product development. Our WorkPlace assessments can be easily contextualized and used for extended workplace applications to solve your most challenging workplace issues and help your organization thrive.

## Take the Next Step

There are many ways to learn what WorkPlace Big Five Profile™ can do for your organization:

- Request a sample report
- Schedule a conversation for detailed answers to your specific questions
- Schedule tryouts for your assessment evaluation team
- Request a predictive validity study for your organization
- Get certified in the WorkPlace Big Five Profile™



## References and Resources

- *Society for Industrial and Organizational Psychology*. The Society for Industrial and Organizational Psychology ([SIOP - The Society for Industrial and Organizational Psychology](http://www.siop.org)) is the premier professional association dedicated to advancing the field of I-O psychology with over 10,000 members worldwide. Helpful articles include the following:
  - *Seven Questions to Ask a Vendor Before Purchasing a Test:*  
<https://www.siop.org/Business-Resources/Employment-Testing/Vendor-Questions>
  - *Types of Employment Tests:*  
<https://www.siop.org/Business-Resources/Employment-Testing/Test-Types>
- *Standards for Educational and Psychological Testing (2014)*. The Standards are a product of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. Published collaboratively by the three organizations since 1966, this document is the premier guidance on testing in the United States and in many other countries.
- *International Test Commission (2013)*. International Guidelines for Test Use ([Intestcom.org](http://Intestcom.org)). The test uses guidelines that relate to the competencies (knowledge, skills, abilities, and other personal characteristics) expected from someone seeking qualification as a test user. Such competencies cover a range of issues including professional and ethical standards in testing, rights of the test taker and other parties involved in the testing process, choice and evaluation of alternative tests, test administration, scoring and interpretation, and report writing and feedback. The guidelines also have implications for standards for test construction, standards for user documentation (e.g., technical and user manuals), and standards for regulating the supply and availability of tests and information about tests.