



Validity and reliability of the WorkPlace Big Five Profile™ 4.0

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 that is estimated to be growing 10% annually. More than 2,500 personality questionnaires are on the market and dozens of new companies appear annually, making the challenge of finding the right assessments all the more difficult.

Our clients consider many factors, and validity and reliability are especially important. Valid and reliable assessments have been proven to be a worthy investment because their solid foundation provides accurate insights that advance business goals. Our assessments also are cost-effective, efficient to implement, and legally sound.

Our clients prioritize psychometrics. They want to know how well the instrument measures what it proposes to measure, and how well it predicts their organization's desired outcomes. This paper is for the decision maker and contains information that expert psychometricians typically request when they evaluate assessments.

The WorkPlace Big Five Profile™ is one of the most psychometrically robust tools on the market. Our personality assessment's average coefficient alpha of .83 is among the highest of all assessments. It complies with all International Test Commission guidelines.

We use current standards

In the early 1950s, researchers focused on three aspects of test validity: content, criterion and construct. Now, validity is viewed as "the degree to which evidence supports intended uses of a



test" (*Standards for Educational and Psychological Testing*, 1999 & 2014). The Society for Industrial and Organizational Psychology updated its *Principles for the Validation and Use of Personnel Selection Procedures* (2003) in order to be consistent with the 1999 Standards. We use the most current standards to ensure that the WorkPlace Big Five Profile™ is valid and reliable.

What does it mean to call a test reliable?

Reliable tests get consistent results when the tests are repeated. Reliability declines over time, however, because experience and context can influence participants' responses to behavioral questions. The Big Five dimensions of Need for Stability (N) and Accommodation (A) are more susceptible to environmental influence than the other three dimensions. It is not surprising that participants' environments and position in the hierarchy affect the patterns in their responses to stress (measured via N) and their level of accommodation (measured via A). The genetic part of trait structure does not change, but the environmentally influenced part can—and does.

Reliability is typically measured by test/retest studies.

- a) **Short-term test/retest.** The test is administered, and then administered again to the same people one to three months later. Well-constructed tests should yield short-term test/retest reliability of around .90. The WorkPlace short-term test/retest reliability of supertraits averages .88.
- b) **Long-term test/retest.** The test is administered, and then administered again to the same people one to three years later. Well-constructed tests should yield long-term test/retest reliability of around .70. The WorkPlace long-term test/retest reliability for supertraits averages .72.

Reliability is also assessed in two other ways: split/half and coefficient alpha. Split/half methods are more common with ability tests, which have right and wrong answers. We have not done



split/half studies with our Big Five tests. Coefficient alpha is a versatile statistic that is accepted as an indicator both of validity and reliability. Good alphas support validity, in that they suggest that all items measure the same thing. They also support reliability, in that they indicate respondents are answering consistently. Alpha values above .7 are generally considered acceptable and satisfactory. Alpha values above .8 are quite good. Alpha values above .9 reflect exceptional internal consistency—perhaps too good, as scores this high can indicate that the items measuring the construct are too similar. The WorkPlace Big Five Profile™ coefficient alpha of .83 is among the highest of many assessments.

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. "Overall IQ" has broad bandwidth, while "3-dimensional spatial rotation ability" has narrow bandwidth. In the WorkPlace, the supertrait Extraversion has broader bandwidth, while the subtrait E:2 sociability has narrower bandwidth. Particularly when using an assessment for selection and coaching, it is important to have both super- and subtraits because some work issues are better explained by broad bandwidth traits (such as achievement being explained by Consolidation), while other, more specific work issues are better explained by narrow bandwidth traits (such as sales achievement being explained by the subtraits of low N:4 rebound time, high E:2 sociability, low A:3 humility, and high C:3 drive, in addition to the supertrait Consolidation). For this reason, the WorkPlace's five supertraits have broader bandwidth and the 23 subtraits have narrower bandwidth.

Compliance with employment law. The original item list for the WorkPlace was more than 800 statements. We asked a labor/employment attorney to note any item we could not ask a prospective employee before hiring, and we eliminated all those items. The WorkPlace has no items that inquire about political, religious, or social beliefs.

Comprehensive vs. partial. Partial scales leave out some aspect of normal personality. For example, the MBTI is partial, as it leaves out Need for Stability. The WorkPlace and other Big Five assessments that are based on the Five-Factor Model are regarded as comprehensive, especially when subtraits are included. The more subtraits, the more comprehensive.

Court record. Of course, one of the best indicators of validity is a) the absence of court challenges, and/or b) the successful defense of court challenges. To our knowledge, no Big Five test has

"gone to court." The WorkPlace has not been challenged.

Cross-validation. A standard analysis for establishing the validity of a test is to divide the norm group randomly in half and apply the scoring algorithms to each group. The results should be the same. We performed a cross-validation study for the WorkPlace, with excellent results.

Empirical vs. theoretical. Tests based on a theory of personality can only be used with that theory of personality, so you have to subscribe to that theory in order to use the test. This is true for the MBTI, AVA, DISC, and so forth. Empirical tests generally try to measure the basic structure of personality, and the results can be used with almost any theory. The Big Five in general (and the WorkPlace in particular) is empirical.

Internal consistency. Each trait is measured by a set of items. For example, the subtrait C2: organization is measured by:

- a) Gets organized before beginning a task (+)
- b) Is neat and tidy (+)
- c) Keeps everything in its place (+)
- d) Organizes for work effectively (+)
- e) Spends time searching for misplaced things (-)

C2 is said to be internally consistent to the degree that respondents tend to answer the first four items similarly, and the last item in the opposite way. This is measured by coefficient alpha, also known as Cronbach's alpha. Alphas should never fall below the value of .5 for subtraits or .7 for supertraits. Low alphas indicate that the items are probably not measuring the same thing. On the other hand, alphas that are too high (over .9), suggest that the items may be too similar and are therefore not robust enough to capture the subtlety and complexity of human behavior.

All WorkPlace subtraits are between .6 and .8, and all supertraits are around .8. This is comparable

to the best assessments available, such as the NEO tests. Interestingly, traits that are more external and easier to observe tend to have higher alphas than traits that are more internal and trickier to observe. An example would be C2: organization (.79) versus N3: interpretation (.62).

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:

- a) **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".
- b) **Global/cultural context.** The WorkPlace has been translated into several other languages. In each case, we translate into the target language, then have a different linguist translate back into English. We "jury" that final translation to ensure that 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 both natural for the target culture(s) to reflect the construct to be 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 should not be used for selection because the results do not allow accurate comparisons of an individual to the population at large. Ipsative scales (e.g., the DISC, AVA, and MBTI) measure individuals against themselves, while normative scales measure individuals against others. Ipsative scales force choices by asking participants to choose which of several options they prefer. They don't reveal whether these choices are strong or mild



preferences. Normative scales ask to what degree participants like something. In doing so, normative scales evaluate one issue at a time, and thus avoid confounding it with other issues. The WorkPlace is a normative assessment, specifically for the work environment.

Norms. Tests should be normed on the same kind of population that will use its results. For example, a test that was normed on a small group (n=150) of managers, mostly male, would not be appropriate (or accurate) to use with female participants, or with other roles, such as sales. Not only must the norm group be similar to the people who will be taking the test, but the norm group must reflect the diversity of that group. The WorkPlace is normed on adults age 18 and older who are working full time. We use the 2009 American Community Survey conducted by the U.S. Census Bureau to help us create a well-balanced norm group. Our initial analysis included more than 60,000 subjects, but the final norm group was reduced to 1,200 so that the number of representatives for each sex, race, age, industry, and job category would reflect their normal distribution in the workforce. The details are available in the *Professional Manual for the WorkPlace Big Five Profile™ 4.0*.

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. Combining traits into a multiple regression (e.g., using the subtraits N4: rebound time, E1: warmth, E2: sociability, E3: activity mode, A3: humility, and C5: methodicalness) to predict the criterion (e.g., sales volume) typically achieves coefficients around .40. As traits are only a part of the total person, we recommend combining trait predictors with other predictors relevant to the performance criterion being measured. Other predictors include mental ability (e.g., numerical analysis), physical ability (e.g., hand-eye coordination), background checks (e.g., credit, police, academic), and experience factors (e.g., military service, previous work experience, hobbies). None of these individual factors associate more than .50 with performance. However, combining them in a multiple regression reaches a coefficient of about .90. In most circumstances, situation-specific validity evidence is



more 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 to select employees leads to the predicted, effective results, such as:

- a) Reductions in employee turnover
- b) Increase in performance levels
- c) Improved information for the acquiring manager to use in coaching and team building
- d) Excellent discriminant validity in one recent study for a state government department.

Each of the six department jobs we analyzed yielded its own unique profile.

Social desirability. Many tests have "lie scales" (also called validity scales or social desirability 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. However, in some groups the tendency to self-enhance (sales folks tend to do this on Accommodation and Consolidation) is so widespread that it makes no difference. The WorkPlace questionnaire includes an honesty pledge. Respondents agree to truthful responses before proceeding with the test. We are completing our analysis on the usefulness of this approach. The

WorkPlace report to consultants alerts them to six response sets that could invalidate the results. A response set is a pattern of responses that suggest the possibility that the respondent is not answering truthfully. For example, the response set "Tendency to agree" is flagged whenever the respondent agrees or strongly agrees with 77 or more of the 107 items used to measure traits. This is because less than 1% (12 out of 1,200) of the norm group agree to that extent.

Want more details about the data?

Here is a quick look at how much data is available. For an expanded review of the psychometrics of the assessment, refer to Section 6 of our *Professional Manual for the WorkPlace Big Five Profile™ 4.0*.

Norm group data. We provide a detailed description of the norm group, including job and industry representation in the norm group, and the means and criteria for selecting our global norm group. Insights can be used to give more precise feedback such as data-driven understanding of global similarities and differences of statistical significance.

Scoring. We share the scoring algorithms for the long and short form, and the subtrait weightings for each of the five supertraits to help you understand the complex calculations behind the user-friendly reporting.

Reliability. We present the coefficient alphas of the long form for each of the five supertraits and 23 subtraits. It lays forth the methodology and reasoning that balances practical utility with statistical soundness.

Validity. Most research personality psychologists regard the NEO PI-R as the current standard of measurement. Based on the principle of validity generalization, we detail the correlation of the WorkPlace 4.0 subtraits with the NEO PI-R factors and facets. With respect to predictive validity, we detail studies that establish the ability of WorkPlace as an effective predictor of future behavior. We also share details of the studies that establish the WorkPlace as having excellent discriminant validity. Using validity generalization, we offer an extensive database that lists the WorkPlace 4.0



means (standard scores) for 16 industries across 16 job roles and 16 job roles across 16 industries.

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
- Employee engagement
- Hiring and selection
- Leadership development
- Performance development
- Personnel selection
- Project team design
- Research and validity studies on work-job fit
- Succession planning
- Team building

Why choose the WorkPlace Big Five Profile™?

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

- It has one of the best sets of alphas.
- It is based purely on the Five-Factor Model, the gold standard of personality assessment, which has been highly validated, widely researched, extensively used, and successfully applied. As you know, this purity enhances reliability.
- It provides accurate insights into complex human behavior. Some assessments lose nuance

by oversimplifying.

- It complies with legal guidelines of the Equal Employment Opportunity Commission and the Americans with Disabilities Act. Some assessments can't legally be used for selection.
- It adheres to International Test Commission guidelines and the *Standards for Educational and Psychological Testing*.
- Paradigm is forward thinking. Ask about our special reports, our other assessments, our products in development, our publications, and our research studies. Ask about how the assessment can solve your workplace issues and support your goals to thrive. Our WorkPlace assessments can be easily contextualized and used for extended workplace applications.

Take the next step

What would you like your next step to be in evaluating the WorkPlace Big Five?

1. Request detailed psychometric information, available in Section 6 of the *Professional Manual for the WorkPlace Big Five Profile™ 4.0*.
2. Request sample reports.
3. Schedule conversation for detailed questions.
4. Schedule tryouts for the assessment evaluation team.
5. Request a predictive validity study for your organization.
6. [Get certified in the WorkPlace Big Five Profile™](#).

Resources and references

- *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 (2001). International Guidelines for Test Use, *International Journal of Testing*, 1(2), 93-114. The test use guidelines relate to the competencies (knowledge, skills, abilities and other personal characteristics) expected from someone seeking qualification as a test user. Such competencies cover such issues as 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.
- *Professional Manual for the Workplace Big Five Profile™ 4.0*.
- [Seven Questions to Ask a Vendor Before Purchasing a Test](#), Society for Industrial and Organizational Psychology.
- [Personality tests – Advantages and Disadvantages](#), Society for Industrial and Organizational Psychology.



ABOUT PARADIGM PERSONALITY LABS

Paradigm Personality Labs is a leader in applying the Five-Factor Model of personality to work and educational settings. Dr. Pierce Howard and Jane Mitchell Howard are the authors of the **WorkPlace Big Five Profile™**, Paradigm's primary assessment.

For more information about Paradigm and the WorkPlace Big Five Profile™, visit www.paradigmpersonality.com or email info@paradigmpersonality.com.

Paradigm Personality Labs
4701 Hedgemore Drive Suite 210
Charlotte NC USA 28209-2200
+1-704-331-0926