

PERSONALYSIS[®]

**PERSONALYSIS
SUMMARY OF
RELIABILITY AND
VALIDITY**

Personalysis White Paper

PURPOSE

This white paper summarizes the recent psychometric studies evaluating Personalysis reliability and validity

Lawrence G Weiss PhD
Test Development and Research
Consultant

Linda K Layton PhD
Organizational Development
Consultant

Contents

INTRODUCTION	2
THE INSTRUMENT	2
RELIABILITY	2
TEST-RETEST	2
<i>Classification Consistency</i>	3
VALIDITY	4
CONSTRUCT VALIDITY	4
TEST INTERPRETATION	5
SUMMARY	6
REFERENCE	7
ABOUT THE AUTHORS	8

Introduction

Psychological tests are scientific instruments carefully developed according to guidelines established by relevant professional associations (APA, AERA, NCME, 2014). As with all psychological assessment instruments, personality tests must be shown to be scientifically reliable and valid for the purpose they are used. Referred to as the *psychometrics* of the test, the supporting data are typically presented in a test manual or technical report. Personalysis has developed a series of such reports to give the test user confidence in the results, but they can be difficult to interpret for those without training in psychometrics. This white paper summarizes in non-technical language the Personalysis psychometric studies conducted to date.

The Instrument

Personalysis is a tool for self-awareness, team awareness and the development action that arises from such awareness. It is used to provide working adults with feedback that helps them understand and describe their behavior at work. Personalysis measures four overarching personality traits. For simplicity, the traits are color coded red, yellow, blue, and green – referred to as Colors. Because people behave differently according to the situation, the four traits are measured within three different motivational contexts, referred to in practice as Dimensions: Preferred, Social, and Instinctive.

The combination of four traits applied to each of the three dimensions produces a unique profile consisting of 12 styles or scales. For each of the 12 scales, any score above 3 points is interpreted to mean the participant easily demonstrates the strengths and tendencies of that trait within its dimension (e.g., Instinctive Red, or Preferred Blue); scores below 3.0 are interpreted to be strengths and tendencies that are less accessible to the user participant. For more information about the history, theory, and structure of the instrument, please visit Technical Report #1.

Reliability

Reliability studies measure the consistency of the test's items, and the dependability of the test's scores over time. Here we review the evaluation of Test-Retest Reliability, and Classification Consistency.

Test-Retest

Test-Retest reliability is the extent to which the results of a test are the same as the results of the same test taken a second time. In this study, 718 working adults completed the test about two weeks apart and their scores on each scale were compared between first and

second testing. An average retest reliability coefficient of .71 and a median coefficient of .74 are encouraging results that indicate a high degree of consistency of test scores between the two time points.

Classification Consistency

For the given test scales, the most critical evidence of reliability is a form of test-retest reliability known as Classification Consistency Reliability (often referred to as Decision Consistency) which evaluates the dependability of the test interpretation and resulting feedback. In other words, it is important that a test participant receive the same interpretation and resulting feedback about their results if they were to take the test again a week or two after taking it the first time, thus demonstrating the consistency of the test across time.

While Test-Retest reliability measures the extent to which the test taker's scores are the same at time one and time two, Classification Consistency analyzes the extent to which participants results and feedback remain the same between time one and time two. In other words, the percentage of time the test participants interpretation and feedback from the results would not change. For example, if a participant is classified as Red Social at time one, how likely are they to be classified as Red Social if they take the test a second time? Strong classification consistency over time gives test users confidence in the dependability of the test results.

The results here were very strong. Across the 12 Personalysis scales, classification consistency ranged from 76% for the Instinctive Red scale to 86% for the Social Red scale with an average of 81% demonstrating that the test produces dependable results. In other words, we would expect, on average, that 81% of participants will receive the same feedback the second time they take the test as they received the first time. Further, only 19% of test takers with scores on the border between two style designations crossed into a different designation when taking the test again. The average difference between scores at times one and two was less than one point for each of the twelve scales, further confirming the dependability of the scores over time.

Sensitivity analysis. A sensitivity analysis was performed to understand the sensitivity of the initial scores at test time one relative to test time two. The data showed that the higher an individual's score at time one, the more likely they were to receive the same style designation when tested again. 87% of those with moderately high scores and 92% of those with very high scores received the same classification when tested a second time.

A detailed review of Test-Retest Reliability and Classification Consistency is evaluated in Technical Report #3: Reliability Estimates of Personalysis Scales.

Validity

Another important form of scientific evidence is validity. Validity studies determine whether each scale measures the trait it intends to measure. In other words, does the test measure what it says it does. One method to assess the test's ability to measure what it says it does is to examine its *construct validity*. This examination allows for confident interpretation of the test's scores.

Construct Validity

A common way of evaluating construct validity is to show correlations between two tests. A pattern of strong correlations with other scales measuring similar constructs is considered evidence of construct validity.

Two research studies were conducted to evaluate construct validity of the Personalysis scales. In the first study, 295 working adults completed both Personalysis and the Myers – Briggs Type Indicator (MBTI). Prior to analyzing the data, 4 experts in personality research formed hypotheses about how each Personalysis scale would be related to specific MBTI scales, based on personality theory. 17 hypotheses were formed. The data showed that all 17 correlations were in the predicted direction, and 14 (82%) were large enough to be meaningful.

Further, 90% of the Personalysis scales tested in this study were supported by one or more strong correlations with MBTI scales. For example, Yellow in the Preferred dimension is highly related to Extroversion on MBTI, and Green in the Social dimension is highly related to the MBTI Judging and Sensing scales. The overall pattern of these relationships supports the theory underlying the Personalysis test.

In a separate research study, construct validity was further evaluated using correlations between Personalysis and the Sixteen Personality Factor Questionnaire (16 PF), another well-established measure of personality. Again, the four experts made 26 predictions between specific Personalysis and 16PF scales based on theory. 423 working adults completed both measures. 24 correlations were in the predicted direction, and 17 were large enough to be meaningful. Again, the pattern of these relationships supported the theory of the test.

Taken together, the results of these two validity studies provide strong support for the construct validity of the Personalysis scales. Overall, 92% of the Personalysis scales were supported by one or more meaningful correlations with either MBTI or 16PF. A detailed review of Validity is evaluated in Technical Reports #4 and #5.

Test Interpretation

Evidence of construct validity informs test interpretation. Strong and predictable correlations between tests measuring similar constructs allows confident interpretation of what the test's scores mean for individuals.

The patterns of correlations with MBTI and 16 PF found in Technical Reports #4 and #5 informs interpretation of each Personalysis scale as follows:

Table 1

Personalysis		MBTI	16 PF
Dimension	Color		
Preferred	Red	None	Assertive
	Yellow	Extroverted	Group Oriented, Warm, Gregarious, Friendly, Fortright,
	Blue	Intuition	Complex, Sensitive
	Green	Introverted, Judging	Orderly
Social	Red	Sensing	None
	Yellow	Extroverted	Warm, Friendly
	Blue	Perceiving, Intuition	Complex, Intellect
	Green	Judging, Sensing	Dutiful, Self-reliant
Instinctive	Red	Thinking	Assertive
	Yellow	None	None
	Blue	Feeling	Self-Assured, Emotional
	Green	Thinking, Judging	Orderly

The magnitude of the above correlations suggests that while Personalysis is partially related to specific MBTI and 16PF scales, it also measures personality traits and styles beyond those instruments. Supported by these validity findings, interpretation and feedback further draws on Personalysis' unique theory to help test takers understand how they prefer to engage, contribute, and communicate when working with others, and how they prefer to think, solve, and decide when faced with work problems.

For example, according to Personalysis theory, Red in the Preferred dimension measures a preference to engage in leadership roles that are goal and action oriented. This individual prefers to achieve results one goal at a time and be personally involved and busy in pursuit of those goals. Further, Yellow in the Instinctive dimension measures a decision-making style that includes a need to understand options, the opinions of others, trends, who else needs to be involved in the decision, and who the decision will impact as critical input into their own decision. There is no corollary in the design of Jungian-based MBTI for Preferred Red nor is there a corollary in the MBTI and 16 PF models for the Instinctive Yellow trait. In these ways, the instrument holds true to a primary goal of its original design – to provide a multifaceted, deeper understanding of individual differences.

Summary

Based on several carefully conducted research studies summarized above, customers can trust the Personalysis test to give them reliable and valid scores about their personality styles as well as those of their teammates and colleagues in work settings.

For further details about these studies the reader is referred to the suite of Personalysis Technical Reports 2-5.

Reference

American Psychological Association, American Educational Research Association & National Council on Measurement in Education, 2014. *The Standards for Educational and Psychological Testing*. Washington, D.C.

About the Authors

Dr. Weiss is former Vice President of Global Research & Development for Pearson Clinical Assessment, the largest developer of psychological tests in the world, where he worked for 27 years prior to becoming a private test development consultant. In his capacity at Pearson, he oversaw the research and development of several hundred widely used psychological tests. He has published 9 textbooks on psychological testing and has been an invited speaker in 13 countries across 5 continents. As an independent consultant to Personalysis Corporation, Dr. Weiss was responsible for designing the research described in this paper and providing expertise and oversight to ensure the quality of the data collection, statistical analyses, and final technical reports.

Dr. Layton is an organizational development consultant and leadership coach with a unique background in organizational psychology, human resources management, business administration, and finance. Her research area of interest is the relationship between personality and communication competence in organizational leaders. As a Personalysis Master Practitioner, she has acquired a depth knowledge about the history of the instrument, performed research utilizing the instrument, and over 20 years of practical application. As an independent consultant to Personalysis for the research described in this paper, Dr. Layton was responsible for program team effectiveness, including collaborative design and implementation, in addition to lending her instrument experience and expertise in the formulation and evaluation of hypotheses.