PERSONALYSIS®

Personalysis Technical Report #4 and #5:

#4: Convergent and Discriminant Validity with the MBTI
#5: Convergent and Discriminant Validity with the 16PF

Authors:
Robert W. Loy, MA
(Industrial and Organizational Psychology),
Loy Consulting

Lawrence G. Weiss, PhD
(Industrial and Organizational Psychology),
Test Development Consultant

Linda Layton, PhD
(Organizational Psychology),
PentaFold Development

January 13th, 2023

Copyright (c) 2023 by Personalysis Corporation, Inc.
675 Bering Drive,
Houston, TX 77057

PERSONALYSIS®

Personalysis Technical Report #4: Convergent and Discriminant Validity with the MBTI

Authors:
Robert W. Loy, MA
(Industrial and Organizational Psychology),
Loy Consulting

Lawrence G. Weiss, PhD
(Industrial and Organizational Psychology),
Test Development Consultant

Linda Layton, PhD
(Organizational Psychology),
PentaFold Development

January 13th, 2023

Copyright (c) 2023 by Personalysis Corporation, Inc.
675 Bering Drive,
Houston, TX 77057

Table of Contents

Abstract

The purpose of the Personalysis construct validation studies is to evaluate the convergent and discriminant validity of Personalysis assessment scales relative to the well-established personality instruments Meyers-Briggs Type Indicator (MBTI) and Sixteen Personality Factor Questionnaire (16PF). This technical report describes the construct validity results of Personalysis and the MBTI. This study was both confirmatory and exploratory. In the confirmatory phase, hypothesized relationships were constructed for 11 of the 12 Personalysis scales with specific MBTI scales. In the exploratory phase, results were analyzed for all remaining MBTI scales. Data were collected on 295 MTurkers who completed both instruments. 10 of the 11 (91%) Personalysis scales had a correlation of .19 or higher with a theoretically related MBTI scale. Further, 10 of the 11 (91%) Personalysis scales demonstrated convergent validity. Overall, the results suggested that ten of the Personalysis scales demonstrated adequate construct validity with the MBTI. The findings of the MBTI validity study are interpreted in context with findings from the 16PF validity study previously reported.

Introduction

In simple terms, construct validity is the degree to which a scale correlates with other scales believed to measure the same or similar constructs (convergent validity), and does not correlate with scales believed to measure different constructs (discriminant validity). The overall pattern of these sets of relationships defines the nomological network within which the scale is situated and provides evidence in support of its construct validity.

In the current study we seek to address the construct validity of the Personalysis assessment by evaluating both the convergent and discriminant validity of its scales using correlations with an established measure of personality. This approach helps to build the nomological network of the Personalysis scales by understanding the relationships of these scales with other measures.

The Standards (AERA Standards, 2014) state validity is a unitary concept where various *aspects* of validity contribute to the inferences made from scale scores. Two sources of construct validity are convergent and discriminant validity. Convergent validity is the relationship between test scores and other measures purporting to measure a similar construct, while discriminant validity is the relationship between test scores and other measures purporting to measure unrelated constructs (AERA Standards, 2013; Cronbach & Meehl, 1955; SIOP Principles, 2018).

We used correlational analyses to estimate the relationship between two scale scores in the current study. A correlation is a measure of the strength of the relationship between two variables (Gatewood et al., 2013). Correlations range from -1.0 to 1.0 (Haldun 2018). The higher a correlation between two scales, the more likely these two scales are related. For instance, correlations above .70 may suggest that two scales are largely redundant. Correlations between .20 and .60 suggest that the two scales measure similar constructs, but each has some unique aspects as well. In contrast, correlations below .20 suggest that the two scales do not measure similar constructs, and thus have discriminant validity.

The current study uses three approaches to evaluate the construct validity of the Personalysis assessment, generating two research questions. First, hypotheses were proposed between Personalysis scales and relevant MBTI scales. Thus, the first research question is presented below:

Research Question 1: Do Personalysis scales correlate higher than or equal to .20 with *hypothesized* MBTI scale scores?

Second, a linear regression was analyzed to build understanding regarding how all hypothesized MBTI scales considered jointly explain each Personalysis scale. It was expected that considering additional MBTI predictors would explain additional variance in Personalysis scales. No research question is proposed for the linear regression analysis, instead, the linear regression analysis is performed to build further understanding of Personalysis scales.

Third, a multitrait-monomethod matrix was computed to assess both convergent and discriminant validity. Convergent validity is the extent to which a scale is related to well established Personality scales. Hypothesized relationships between Personalysis and MBTI scales were used as evidence of convergent validity in the current study. Discriminant validity is the degree to which Personalysis scales are related to theoretically unrelated scales. Furthermore, discriminant validity is used in comparison to convergent validity estimates, such that convergent validity estimates should be higher than that of discriminant validity.

A multitrait-monomethod matrix provides evidence of convergent and discriminant validity by comparing hypothesized to non-hypothesized correlations. The purpose of the multitrait-monomethod matrix is to evaluate if Personalysis scales correlate higher with scales they *should* than those they *shouldn't*. The following terms are used are used to help the reader follow the construct validity terminology:

a) *Hypothesized scales (convergent validity evidence)* refer to MBTI scales that are believed to be theoretically linked to the Personalysis scale based on a priori judgements of four Personalysis experts, providing evidence of convergent validity. See Personalysis Technical Report #2: Methodology of Personalysis Studies for details concerning the experts experience and the procedure for developing the hypothesized relationships.

b) *Non-hypothesized scales (discriminant validity evidence)* refer to MBTI scales that were not believed to be theoretically linked to the Personalysis scales according to expert judgment, providing evidence of discriminant validity.

A significant portion of the current study rests upon the multitrait-monomethod matrix approach. Below are the specific research questions we posed:

Research Question 2: Do Personalysis scales correlate higher with *hypothesized* MBTI scales than *non-hypothesized* MBTI scales?

A focus of the current study is to assess the overall pattern formed by a set of relationships between multiple scales, thus creating a nomological network of relationships. That is, assessing how the purported constructs measured by scale scores are related to the other hypothesized relationships (Cronbach & Meehl, 1955; Smith, 2005). A nomological network is an "interlocking system of laws which constitute a theory" (Cronbach & Meehl, 1955, pg. 11). Nomological networks are the basis for theoretically understanding what the construct validity of an assessment is and provide support (or refute) inferences made from assessment results. The purpose of this study is to evaluate validity evidence for each Personalysis scale and develop understanding of each Personalysis scale's nomological network.

Although we made specific hypotheses, we do not necessarily consider an unsupported hypothesis as discrediting the construct validity of the assessment. Instead, we are seeking to improve our understanding of how these theoretically derived scales compare to other established personality scales. Consider a small observed relationship between two scales that were hypothesized to be highly related. Such findings improve understanding of what this scale is *not* rather than what it is. Similarly, an unanticipated but strong observed relationship between two scales adds to the understanding of what the scale is measuring. Thus, unsupported hypotheses and post-hoc analysis can provide evidence of validity that is useful in better understanding a construct, even if not predicted.

Method

Sample

The sample consisted of 295 participants selected to represent the population of working adults by gender, ethnicity, race, age, and educational level. The data were collected using MTurk, a widely used crowd sourcing tool. Data collection procedures are described in detail in Personalysis Technical Report #2 including inclusion/exclusion criteria, data quality controls, and data cleaning

procedures. The demographics of the final sample provide a reasonable representation of the Personalysis target population. Study demographics are as follows.

Table 1

Sample Demographics

Gender	*n*	%	Census estimates
Female	118	40%	47%
Male	167	57%	53%
Other*	10	3%	n/a
Ethnicity			
Hispanic	38	13%	12%
Non-Hispanic	256	87%	88%
Race			
Another race	2	1%	n/a
Asian or Asian American	21	8%	8%
Black or African American	30	12%	12%
Two or more races	13	5%	n/a
White or Caucasian	194	75%	77%
Age			
20-24	23	8%	9%
25-54	225	76%	66%
55-64	32	11%	18%
over 65	15	5%	7%
Educational Attainment			
Less than high school or other	1	0%	8%
High school graduates	1	0%	23%
Some college, associate's degree, or vocational training	111	38%	29%
Bachelor's degree or higher	182	62%	41%

Notes. *Other denotes non-binary/non-conforming (*n* = 8) and transgender man (*n* = 2). *N* = 295.

Instruments

All participants completed both the Personalysis and MBTI. The characteristics of these instruments are described in Personalysis Technical Report #2: Methodology of Personalysis Studies.

Procedures

All analyses were conducted using the R Statistical Software (v4.2.2; R Core Team 2022). There were three approaches to conduct the following analysis. First, a bivariate approach was used to examine the relationship between Personalysis scales with hypothesized MBTI scales. Second, a multivariate analysis was conducted running linear regressions using all hypothesized MBTI scales regressed onto the targeted Personalysis scale. The package "stats" (v4.2.2; R Core Team 2022) was used to conduct the multivariate analysis. Third, a bivariate analysis was conducted to compute a correlation matrix of all Personalysis scales with their highest hypothesized MBTI scale and all non-hypothesized MBTI scales. Correlations within each of these three categories were averaged and presented as a multitrait-monomethod matrix.

RESULTS

Investigation of Hypothesized Relationships Between MBTI Scales Using a Bivariate Correlational Analysis

The current section reports the correlation between each Personalysis scale and the hypothesized MBTI scale(s). There were two methods to evaluate the hypothesized relationship: 1) is the hypothesized relationship in the expected direction; 2) is the hypothesized relationship larger than or equal to .20. A threshold of .20 was used to establish practical significance (Kirk, 1996).

As stated in Personalysis Technical Report #2, a team of four experts derived hypotheses for each of the 12 Personalysis scales, with the exception of the Instinctive Yellow scale for which no relationships with MBTI scales were predicted. There was at least one hypothesized relationship for each of the remaining 11 Personalysis scales. There were a total of 17 hypothesized relationships between Personalysis and MBTI scales. All 17 relationships were in the expected direction. Further, 14 of the 17 (82%) hypothesized relationships were supported using the .20 threshold, or marginally supported at .19 (Social Red with MBTI Sensing).

Red Scales

Regarding the three Red scales, correlations range from .08 to .35 with the hypothesized MBTI scales. All three of these relationships were in the expected direction. One of the three hypothesized relationships was supported according to the .20 hypothesis cutoff, and the Social Red with MBTI Sensing hypothesized relationship was marginally supported at .19.

Yellow Scales

Two relationships were hypothesized between Yellow Personalysis and MBTI scales. Both Preferred Yellow and Social Yellow were in the prediction direction and the hypotheses were supported.

Blue Scales

There were a total of six Blue scale hypotheses. All six of these relationships were in a positive direction, correlations ranging from .02 to .71. Five of the seven (71%) hypothesized relationships were supported. Additionally, all three Blue scales had at least one hypothesized relationship that was supported. For Preferred Blue, both hypotheses were in the expected direction, and the hypothesis with Intuition was supported. All three hypotheses proposed for the Social Blue

were in the expected direction, of which two of the three (67%) were supported at the .20 level. The average correlation of the two supported hypotheses was .45. A single hypothesis was proposed for the Instinctive Blue scale was supported and in the expected direction.

Green Scales

A total of six Green hypothesized relationships were observed. All six (100%) of the hypothesized relationships were in the predicted direction and supported, ranging from .20 to .62. For the Preferred Green scale, both hypotheses were in the expected direction and supported at the .20 level. The average of these supported relationships was .32. For the Social Green scale, both hypotheses were in the expected direction and supported at the .20 level. The average of these supported relationships was .49. Lastly, Instinctive Green had two hypotheses proposed, of which both were in the predicted direction and supported at the .20 level.

Overall, 10 of the 11 Personalysis scales tested, or 91%, were found to be supported by a correlation of .19 or higher with at least one hypothesized MBTI scale. Only the Preferred Red scale was unsupported by any hypothesized MBTI scale. Post hoc analyses revealed a .45 correlation between Preferred Red and MBTI Thinking.

Table 2
Hypothesized Correlations Between Personalysis and MBTI Scales

			Hypothesized	
Red	MBTI	Correlation	Direction Supported	Relationship supported
Preferred	Extroversion	.08	Yes	No
Social	Sensing	.19	Yes	Marginally
Instinctive	Thinking	.35	Yes	Yes
Yellow				
Preferred	Extroversion	.34	Yes	Yes
Social	Extroversion	.37	Yes	Yes
Instinctive	–	–	–	–
Blue				
Preferred	Introversion	.13	Yes	No
Preferred	Intuition	.71	Yes	Yes
Social	Thinking	.02	Yes	No
Social	Perceiving	.26	Yes	Yes
Social	Intuition	.64	Yes	Yes
Instinctive	Feeling	.40	Yes	Yes
Green				
Preferred	Introversion	.26	Yes	Yes
Preferred	Judging	.38	Yes	Yes
Social	Judging	.35	Yes	Yes
Social	Sensing	.62	Yes	Yes
Instinctive	Thinking	.21	Yes	Yes
Instinctive	Judging	.20	Yes	Yes
Note. *N* = 295.		Yes %	100%	76%

Linear Regression Analysis

This section examines the relationship between each Personalysis scale and all theoretically predicted MBTI scales simultaneously. This procedure removes overlapping variance common across the unique relationships shown in Table 3, providing a more refined measure of the strength of the relationship between Personalysis and MBTI scales considered jointly. Regression coefficients (β) are reported for each linear regression to indicate the degree to which each MBTI scale contributed unique variance to the hypothesized the Personalysis scale. Furthermore, variance explained is computed by squaring the adjusted Multiple R.

This statistic quantifies the amount of variance in each Personalysis scale that is explained by its theoretically related MBTI scales, considered jointly.

There were five Personalysis scales with more than one hypothesis proposed: two Blue and three Green scales. As shown in Table 3, these five scales had Multiple Rs ranging from .25 to .72.

Blue Scales with Multiple Hypotheses

The Preferred Blue scale demonstrated a Multiple R of .72 with the hypothesized Introversion and Intuition scale. These scales jointly explained 52% of the variance in the Preferred Blue scale. A majority of the variance explained is due to the Intuition scale (β = .13). Moreover, Introversion does explain some of the variance in Preferred Blue (β = .03).

The Social Blue scale demonstrated an adjusted Multiple R of .69 with the MBTI scales Thinking, Perceiving and Intuition. These scales jointly explained 48% of the variance in the Social Blue scale. A majority of the variance explained is due to the Intuition scale (β = .13) while Thinking also explains some variance in the Social Blue scale (β = .06).

These findings suggests that the hypothesized relationships contribute meaningful variance to the Preferred Blue and Social Blue scales.

Green Scales with Multiple Hypotheses

Preferred Green had an adjusted Multiple R of .43 with Introversion and Judging. These scales jointly explained 18% of the variance in the Preferred Green scale. Both the Introversion (β = .09) and Judging (β = .05) scales explain a meaningful amount of variance in the Preferred Green scale.

Social Green demonstrated an adjusted Multiple R of .62 with MBTI Judging and Sensing. These scales jointly explained 38% of the variance in the Social Green scale. A majority of the variance explained is due to the Sensing scale (β = .10). Further, Judging explained a small amount of the variance in Social Green (β = .02).

Instinctive Green had an adjusted Multiple R of .25 with Thinking and Judging. These scales jointly explained 6% of the variance in the Instinctive Green scale. Both the Thinking and Judging scales (β = .03) explained an equal amount of variance in Instinctive Green.

These findings suggests that the hypothesized relationships contribute substantial variance to the Preferred Green and Social Green scales, and a smaller but meaningful amount of variance to the Instinctive Green scale.

Table 3
Linear Regression Analysis of Hypothesized Relationships Between Personalysis and MBTI Scales

Blue	Scales	β	Multiple R	Variance explained
Rational	Extroversion	-.03	.72	52%
	Sensing	-.13		
Social	Thinking	.06	.69	48%
	Judging	.00		
	Sensing	-.13		
Green				
Rational	Extroversion	-.05	.43	18%
	Judging	.09		
Social	Judging	.02	.62	38%
	Sensing	.10		
Instinctive	Thinking	.03	.25	6%
	Judging	.03		

Note. $N = 295$.

Multitrait-Monomethod Matrix Analysis

A multitrait-monomethod matrix was computed to further evaluate the construct validity of the Personalysis assessment with the MBTI. In Table 4, there are two statistics used to demonstrate convergent and discriminant validity organized by Personalysis scale. The first statistic, hypothesized, is the largest correlation observed between each Personalysis scale and the theoretically related MBTI scale(s). When more than one hypothesis was made for a scale, the largest univariate correlation was used as this relationship accounts for the most variance in that Personalysis scale. When only one hypothesis was made for a scale, that univariate correlation was entered into the multitrait-monomethod matrix. Second, the average of the non-hypothesized correlations is reported to demonstrate discriminant validity with theoretically unrelated scales. Evidence of convergent and discriminant validity exists when hypothesized correlations are *higher* than non-hypothesized scales.

Red Scales

For the Preferred Red scale, the hypothesized correlation of .08 with Extroversion was lower than the non-hypothesized average correlation ($\bar{r}$ = .23). These findings provide little support for the Preferred Red scale, suggesting the MBTI Extroversion scale is not strongly related to the Preferred Red scale, but that one or more other MBTI scales is related to it.

The Social Red scale demonstrated a .19 hypothesized relationship which is higher than the .10 non-hypothesized average correlation, demonstrating support for the construct validity evidence for the Social Red scale.

The Instinctive Red scale demonstrated a strong hypothesized relationship (r = .35) with the Thinking scale. The Instinctive Red correlation was higher than the non-hypothesized average correlation ($\bar{r}$ = .10), demonstrating strong construct validity evidence.

In sum, strong evidence of construct validity was found for the Social Red, and Instinctive Red scales, and little support for the Preferred Red scale.

Yellow Scales

Preferred Yellow had a .34 correlation with the Extroversion scale. This correlation is higher than the non-hypothesized average correlation ($\bar{r}$ = .27), providing construct validity evidence for the Preferred Yellow Scale.

Social Yellow had a .37 correlation with the Extroversion scale. This correlation is higher than the .19 non-hypothesized average correlation observed, providing construct validity evidence for the Social Yellow scales.

There was no hypothesized relationship for the Instinctive Yellow scale.

In sum, strong evidence of construct validity was found for the Preferred and Social Yellow scales.

Blue Scales

The Preferred Blue scale demonstrated a .71 correlation with the Intuition scale. This value is higher than the .25 non-hypothesized average correlation, thus strongly supporting validity evidence for this scale.

The Social Blue scale had a .64 correlation with the Intuition scale and an average .20 non-hypothesized correlation. Thus, strong validity evidence is found for the Social Blue scale.

Lastly, the Instinctive Blue scale had a .40 correlation with the Feeling scale which is higher than the .11 correlation average non-hypothesized correlation. Thus, strong validity evidence was found for the Instinctive Blue scale.

In sum, strong convergent validity evidence was found for the Preferred, Social, and Instinctive Blue scales.

Green Scales

Preferred Green demonstrated a .38 correlation with the Judging scale, which is lower than the non-hypothesized (.48) correlations. Of importance, post hoc analyses revealed that the non-hypothesized average correlation drops to .35 if Sensing (r = .60) is removed from the analysis, suggesting an unanticipated relationship between Preferred Green and Sensing. Thus, some validity support is found for the Preferred Green scale once Sensing is considered.

Social Green demonstrated a .62 correlation with the Sensing scale as predicted, and an average .20 non-hypothesized correlation. Thus, strong validity evidence was found for the Social Green scale.

Lastly, the Instinctive Green scale demonstrated a .21 correlation with the Thinking scale. This value is higher than the non-hypothesized average correlation ($\bar{r}$ = .17). Thus, validity evidence was found for the Instinctive Green scale.

In sum, evidence of construct validity is observed for the Social and Instinctive Green scales, and some evidence is observed for the Preferred Green scale.

Table 4
Personalysis Scale Correlations With MBTI Scales

Red	Hypothesized*	Non-hypothesized
Preferred	.08	.23
Social	.19	.10
Instinctive	.35	.10
Yellow		
Preferred	.34	.27
Social	.37	.19
Instinctive	–	.18
Blue		
Preferred	.71	.25
Social	.64	.20
Instinctive	.40	.11
Green		
Preferred	.38	.48
Social	.62	.20
Instinctive	.21	.17

Notes. *N* = 295. * Denotes largest hypothesized correlation.

Exploratory Findings

In the exploratory phase, large non-hypothesized correlations were examined between Personalysis and MBTI scales. Because these analyses are exploratory, a higher threshold of $r \geq .30$ was used. As shown in Table 5, post hoc analyses revealed that the MBTI Thinking scale demonstrated meaningful correlations with Preferred Red (r = .45) and Preferred Green (r = .35). The MBTI Feeling scale was correlated with Preferred Yellow (r = .50) and Social (r = .41). Further, MBTI Sensing was correlated with Preferred Green.

Table 5
Exploratory Correlations Larger Than .29
Between Personalysis and MBTI Scales

Red	Scales	Correlation
Preferred	Thinking	.45
Yellow		
Preferred	Feeling	.50
Social	Feeling	.41
Green		
Preferred	Thinking	.35
Preferred	Sensing	.60

Note. N = 295.

Discussion

The construct validity of the Personalysis scales was investigated using several methods. First, each scale was evaluated for convergent validity by examining their relationships to theoretically linked MBTI scales. 17 a priori hypotheses were developed across 11 of the 12 Personalysis scales. 13 of 17, or 76%, of the hypotheses were supported by univariate correlations exceeding the established threshold of .20. Further, 10 of the 11 Personalysis scales tested, or 91%, were found to be supported by a correlation of .19 or higher with at least one hypothesized MBTI scale.

When more than one MBTI scale was hypothesized to relate to a single Personalysis scale, a series of linear regressions revealed that the median amount of variance jointly explained by the theoretically predicted scales was 38% (range 6% to 52%). Substantial variance ($\geq$ 18%) was explained for four of the five multivariate predictions. The MBTI explained a smaller amount of the variance in the Instinctive Green scale.

A multitrait-monomethod matrix showed that 9 of the 11 (82%) a priori hypotheses were further supported by higher correlations with hypothesized than non-hypothesized MBTI scales. The Preferred Red and Green scales showed lower hypothesized than non-hypothesized correlations. Post hoc exploratory analyses revealed that elements of the MBTI Thinking dimension are related to both the Preferred Red and Green scales, and that the MBTI Sensing dimension is related to the Preferred Green scale. Further exploratory analyses revealed that elements of

the MBTI Feeling dimension are related to the Preferred Yellow and Social Yellow scales. Although unanticipated, these post hoc findings are consistent with Personalysis theory, and serve to better articulate it's nomological network. The Preferred Red and Green styles tend to emphasize logic, Preferred Yellow and Social Yellow styles focus on emotional tone, while Preferred Green seeks structure.

In conclusion, the present study found strong convergent validity for 9 of the 12 Personalysis scales. Although convergent validity was unsupported for the Preferred Red scale in the current study, Preferred Red was supported by evidence of convergent validity with the 16PF. Likewise, Preferred Green demonstrated some support in both the current study and the 16PF. There were no hypothesized relationships for Instinctive Yellow in the current study, but some support was found the when considering the 16PF. Further research may explore which scales from previously established personality assessments are most related to Instinctive Yellow and Preferred Green.

Evaluation of a test's validity is an ongoing process involving a collection of different lines of evidence across multiple studies. The results of this study provide strong support for a majority of the Personalysis scales. The results presented here should be interpreted in context with results of the previously reported validity study with 16PF (see Table 6 in Appendix). Taken together, these two studies provide some convergent validity evidence for 11 of the 12 Personalysis scales.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Gatewood, R. D., Field, H. S., & Barrick, M. (2013). *Human Resource Selection* (7th Ed.). Cengage Learning.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.

Haldun, A. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(3), 91–93. https://doi:10.1016/j.tjem.2018.08.001

Myers, I. B. & Briggs, K. C. (1998). *The Myers-Briggs Type Indicator Form M*. Consulting Psychologist Press.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sackett, P. R., Tippins, N. T., Arthur, W., Jr, Delany, T., Dunleavy, E. M., Hayes, T. L., Hough, L., Oswald, F., Putka, D. J., Ryan, A. M., & Schmitt, N. (2018). *Principles for the Validation and Use of Personnel Selection Procedures*. American Psychological Association.

Smith, G. T. (2005). On Construct Validity: Issues of Method and Measurement. *Psychological Assessment, 17*(4), 396–408. https://doi.org/10.1037/1040-3590.17.4.396

Appendix

Table 6

Convergent Validity Support for each Personalysis Scale

Red	MBTI	16PF
Preferred	Unsupported	Supported
Social	Supported	Unsupported
Instinctive	Supported	Supported
Yellow		
Preferred	Supported	Supported
Social	Supported	Supported
Instinctive	–	Some support
Blue		
Preferred	Supported	Supported
Social	Supported	Supported
Instinctive	Supported	Supported
Green		
Preferred	Some support	Some support
Social	Supported	Supported
Instinctive	Supported	Supported

Notes. MBTI *N* = 295. 16PF *N* = 423.

PERSONALYSIS®

Personalysis Technical Report #5: Convergent and Discriminant Validity with the 16PF

Authors:
Robert W. Loy, MA
(Industrial and Organizational Psychology),
Loy Consulting

Lawrence G. Weiss, PhD
(Industrial and Organizational Psychology),
Test Development Consultant

Linda Layton, PhD
(Organizational Psychology),
PentaFold Development

January 13th, 2023

Copyright (c) 2023 by Personalysis Corporation, Inc.
675 Bering Drive,
Houston, TX 77057

Table of Contents

Abstract

The purpose of the Personalysis construct validation studies is to evaluate the convergent and discriminant validity of Personalysis assessment scales relative to the well-established personality instruments Sixteen Personality Factor Questionnaire (16PF) and Meyers-Briggs Type Indicator (MBTI). This technical report describes the construct validity results of Personalysis and the 16PF. This study was both confirmatory and exploratory. In the confirmatory phase, hypothesized relationships were constructed for all 12 Personalysis scales with specific 16PF scales. In the exploratory phase, results were analyzed for all remaining 16PF scales. Data were collected on 423 MTurkers who completed both instruments. 10 of the 12 (83%) Personalysis scales had a correlation of .18 or higher with a theoretically related 16PF scale. Further, 9 of the 12 (75%) Personalysis scales demonstrated convergent validity. Overall, the results suggested that nine of the Personalysis scales demonstrated adequate construct validity with the 16PF. The findings of the 16PF validity study are interpreted in context with findings from the MBTI validity study previously reported.

Introduction

In simple terms, construct validity is the degree to which a scale correlates with other scales believed to measure the same or similar constructs (convergent validity), and does not correlate with scales believed to measure different constructs (discriminant validity). The overall pattern of these sets of relationships defines the nomological network within which the scale is situated and provides evidence in support of its construct validity.

In the current study we seek to address the construct validity of the Personalysis assessment by evaluating both the convergent and discriminant validity of its scales using correlations with an established measure of personality. This approach helps to build the nomological network of the Personalysis scales by understanding the relationships of these scales with other measures.

The Standards (AERA Standards, 2014) state validity is a unitary concept where various *aspects* of validity contribute to the inferences made from scale scores. Two sources of construct validity are convergent and discriminant validity. Convergent validity is the relationship between test scores and other measures purporting to measure a similar construct, while discriminant validity is the relationship between test scores and other measures purporting to measure unrelated constructs (AERA Standards, 2013; Cronbach & Meehl, 1955; SIOP Principles, 2018).

We used correlational analyses to estimate the relationship between two scale scores in the current study. A correlation is a measure of the strength of the relationship between two variables (Gatewood et al., 2013). Correlations range from -1.0 to 1.0 (Haldun, 2018). The higher a correlation between two scales, the more likely these two scales are related. For instance, correlations above .70 and higher suggest that two scales are largely redundant. Correlations between .20 and .60 suggest that the two scales measure similar constructs, but each has some unique aspects as well. In contrast, correlations below .20 suggest that the two scales do not measure similar constructs, and thus have discriminant validity.

The current study uses three approaches to evaluate the construct validity of the Personalysis assessment, generating two research questions. First, hypotheses were proposed between Personalysis scales and relevant 16PF scales. Thus, the first research question is presented below:

Research Question 1: Do Personalysis scales correlate higher than or equal to .20 with *hypothesized* 16PF scale scores?

Second, a linear regression was analyzed to build understanding regarding how all hypothesized 16PF scales considered jointly explain each Personalysis scale. It was expected that considering additional 16PF predictors would explain additional variance in Personalysis scales. No research question is proposed for the linear regression analysis, instead, the linear regression analysis is performed to build further understanding of Personalysis scales.

Third, a multitrait-multimethod matrix was computed to assess both convergent and discriminant validity. Convergent validity is the extent to which a scale is related to well established Personality scales. Hypothesized relationships between Personalysis and 16PF scales were used as evidence of convergent validity in the current study. Discriminant validity is the degree to which Personalysis scales are related to theoretically unrelated scales. Furthermore, discriminant validity is used in comparison to convergent validity estimates, such that convergent validity estimates should be higher than that of discriminant validity.

A multitrait-monomethod matrix provides evidence of convergent and discriminant validity by comparing hypothesized to non-hypothesized correlations. The purpose of the multitrait-monomethod matrix is to evaluate if Personalysis scales correlate higher with scales they *should* than those they *shouldn't*. The following terms are used are used to help the reader follow the construct validity terminology:

a) *Hypothesized scales (convergent validity evidence)* refer to 16PF scales that are believed to be theoretically linked to the Personalysis scale based on a priori judgements of four Personalysis experts, providing evidence of convergent validity. See Personalysis Technical Report #2: Methodology of Personalysis Studies for details concerning the experts experience and the procedure for developing the hypothesized relationships.

b) *Non-hypothesized scales (discriminant validity evidence)* refer to 16PF scales that were not believed to be theoretically linked to the Personalysis scales according to expert judgment, providing evidence of discriminant validity.

A significant portion of the current study rests upon the multitrait-multimethod matrix approach. Below are the specific research questions we posed:

Research Question 2: Do Personalysis scales correlate higher with *hypothesized* 16PF scales than *non-hypothesized* 16PF scales?

A focus of the current study is to assess the overall pattern formed by a set of relationships between multiple scales, thus creating a nomological network of relationships. That is, assessing how the purported constructs measured by scale scores are related to the other hypothesized relationships (Cronbach & Meehl, 1955; Smith, 2005). A nomological network is an "interlocking system of laws which constitute a theory" (Cronbach & Meehl, 1955, pg. 11). Nomological networks are the basis for theoretically understanding what the construct validity of an assessment is, and provide support (or refute) inferences made from assessment results. The purpose of this study is to evaluate validity evidence for each Personalysis scale, and develop understanding of each Personalysis scale's nomological network.

Although we made specific hypotheses, we do not necessarily consider an unsupported hypothesis as discrediting the construct validity of the assessment. Instead, we are seeking to improve our understanding of how these theoretically derived scales compare to other established personality scales. Consider a small observed relationship between two scales that were hypothesized to be highly related. Such findings improve understanding of what this scale is *not* rather than what it is. Similarly, an unanticipated but strong observed relationship between two scales adds to the understanding of what the scale is measuring. Thus, unsupported hypotheses and post-hoc analysis can provide evidence of validity that is useful in better understanding a construct, even if not predicted.

Method

Sample

The sample consisted of 423 participants selected to represent the population of working adults by gender, ethnicity, race, age, and educational level. The data were collected using MTurk, a widely used crowd sourcing tool. Data collection procedures are described in detail in Personalysis Technical Report #2

including inclusion/exclusion criteria, data quality controls, and data cleaning procedures. The demographics of the final sample provide a reasonable representation of the Personalysis target population. Study demographics are as follows.

Table 1

Sample Demographics

Gender	*n*	%	Census estimates
Female	214	50%	47%
Male	203	48%	53%
Other*	6	1%	n/a
Ethnicity			
Hispanic	45	11%	12%
Non-Hispanic	378	89%	88%
Race			
American Indian or Alaska Native	2	0%	n/a
Another race	1	0%	n/a
Asian or Asian American	37	9%	8%
Black or African American	48	11%	12%
Two or more races	14	3%	n/a
White or Caucasian	282	67%	77%
Age			
20-24	19	4%	9%
25-54	329	78%	66%
55-64	56	13%	18%
over 65	19	4%	7%
Educational Attainment			
Less than high school or other	3	1%	8%
High school graduates	43	10%	23%
Some college, associate's degree, or vocational training	139	33%	29%
Bachelor's degree or higher	238	56%	41%

Note. *Other denotes non-binary/non-conforming (*n* = 4), transgender man (*n* = 1), and transgender woman (*n* = 1). *N* = 423.

Instruments

All participants completed both the Personalysis and 16PF. The characteristics of these instruments are described in Personalysis Technical Report #2: Methodology of Personalysis Studies.

Procedures

All analyses were conducted using the R Statistical Software (v4.2.2; R Core Team 2022). There were three approaches to conduct the following analysis. First, a bivariate approach was used to examine the relationship between Personalysis scales with hypothesized 16PF scales. Second, a multivariate analysis was conducted running linear regressions using all hypothesized 16PF scales regressed onto the targeted Personalysis scale. The package "stats" (v4.2.2; R Core Team 2022) was used to conduct the multivariate analysis. Third, a bivariate analysis was conducted to compute a correlation matrix of all Personalysis scales with their highest hypothesized 16PF scale and all non-hypothesized 16PF scales. Correlations within each of these three categories were averaged and presented as a multitrait-multimethod matrix.

RESULTS

Investigation of Hypothesized Relationships Between 16PF Scales Using a Bivariate Correlational Analysis

The current section reports the correlation between each Personalysis scale and the hypothesized 16PF scale(s). There were two methods to evaluate the hypothesized relationship: 1) is the hypothesized relationship in the expected direction; 2) is the hypothesized relationship larger than .20. A threshold of .20 was used to establish practical significance (Kirk, 1996).

As stated in Personalysis Technical Report #2, a team of four experts derived hypotheses for each of the 12 Personalysis scales. There were a total of 26 hypothesized relationships between Personalysis and 16PF scales. 25 of the 26 relationships were in the expected direction. Further, 17 of the 26 (65%) hypothesized relationships were supported using the .20 threshold, and two hypotheses were marginally supported at .18 (Preferred Green with 16PF Orderliness) and .19 (Instinctive Green with 16PF Orderliness).

Red Scales

Regarding the three Red scales, correlations range from -.03 to .42 with the hypothesized 16PF scales. Two of the three relationships were in the expected direction. Both the Preferred Red and Instinctive Red scales were supported according to the .20 hypothesis threshold. The Social Red hypothesized relationship was unsupported.

Yellow Scales

Eight relationships were hypothesized between Yellow Personalysis and 16PF scales. These correlations ranged from .10 to .50. There were five hypotheses for the Preferred Yellow scale, which were all supported in the correct direction. The average of the five hypothesized correlations was .41. Two hypotheses were proposed for the Social Yellow scale, with both hypotheses in the correct direction and supported. The average of these two supported hypotheses was .41. Lastly, Instinctive Yellow scale had a hypothesized correlation of .10, which was in the correct direction, but below the .20 threshold for convergent validity support. Although this .10 correlation is not meaningful, it is the second largest positive correlation between Instinctive Yellow and a 16PF scale. The largest relationship observed is .11 with Gregariousness.

Blue Scales

There were a total of eight Blue scale hypotheses. All eight of the hypothesized relationships were in the predicted direction. These correlations range from .04 to .52. Six of the eight (75%) hypothesized relationships were supported at the .20 correlation level. For Preferred Blue, both hypotheses were in the expected direction and supported. The average of these hypotheses was .38. Both Social Blue hypothesized relationships were in the expected direction and supported. The average correlation of these two supported hypotheses was .43. Three of the four (75%) hypotheses proposed for the Instinctive Blue scale were in the expected direction. Further, two of the Instinctive Blue hypotheses were supported, and they had an average correlation of .44.

Green Scales

A total of seven Green hypothesized relationships were observed. All seven (100%) of the hypothesized relationships were in the predicted direction, ranging from -.07 to .31. Four of the seven (57%) relationships were at least marginally supported. For the Preferred Green scale, one hypothesis was in the expected direction and marginally supported at the .19 level. Two hypotheses were proposed for the Social Green scale, both relationships were in the expected direction and supported with an average hypothesized correlation of .29. Lastly, Instinctive Green had four hypotheses proposed, of which all four were in the predicted direction, but only one hypothesized relationship was marginally supported at the .18 level.

Overall, 10 of the 12, or 83%, of Personalysis scales were found to be supported by a correlation of .18 or higher with at least one hypothesized 16PF scale. Only the Social Red and Instinctive Yellow scales were unsupported by any hypothesized 16PF scale.

Table 2
Hypothesized Correlations Between Personalysis and 16PF Scales

Red	16PF Scales	Hypothesized Direction	Correlation	Hypothesized Direction Supported	Hypothesized Relationship supported
Preferred	Assertiveness	+	.30	Yes	Yes
Social	Assertiveness	+	-.03	No	No
Instinctive	Assertiveness	+	.42	Yes	Yes
Yellow					
Preferred	Warmth	+	.46	Yes	Yes
	Gregariousness	+	.40	Yes	Yes
	Friendliness	+	.36	Yes	Yes
	Forthright	+	.34	Yes	Yes
	Group-oriented	+	.50	Yes	Yes
Social	Warmth	+	.46	Yes	Yes
	Friendliness	+	.36	Yes	Yes
Instinctive	Group-oriented	+	.10	Yes	No
Blue					
Preferred	Sensitivity	+	.25	Yes	Yes
	Complexity	+	.50	Yes	Yes
Social	Intellect	+	.34	Yes	Yes
	Complexity	+	.52	Yes	Yes
Instinctive	Warmth	+	.04	Yes	No
	Emotional Stability	–	.39	Yes	Yes
	Dutifulness	+	.16	Yes	No
	Self-assured	–	.48	Yes	Yes
Green					
Preferred	Orderliness	+	.19	Yes	Marginally
Social	Dutifulness	+	.31	Yes	Yes
	Group-oriented	–	-.27	Yes	Yes
Instinctive	Gregariousness	–	-.07	Yes	No
	Imagination	–	-.15	Yes	No
	Group-oriented	–	-.07	Yes	No
	Orderliness	+	.18	Yes	Marginally
Note. *N* = 423.					73%
			Yes %	96%	

Linear Regression Analysis

This section examines the relationship between each Personalysis scale and all theoretically predicted 16PF scales simultaneously. This procedure removes overlapping variance common across the unique relationships shown in Table 3, providing a more refined measure of the strength of the relationship between Personalysis and 16PF scales considered jointly. Regression coefficients (β) are reported for each linear regression to indicate the degree to which each 16PF scale contributed unique variance to the hypothesized the Personalysis scale. Furthermore, variance explained is computed by squaring the adjusted multiple R. This statistic quantifies the amount of variance in each Personalysis scale that is explained by its theoretically related 16PF scales, considered jointly.

There were seven Personalysis scales with more than one hypothesis proposed: two Yellow, three Blue, and two Green scales. As shown in Table 3, these five scales had Multiple Rs ranging from .21 to .56.

Yellow Scales with Multiple Hypotheses

Preferred Yellow demonstrated a Multiple R of .56 with the five hypothesized 16PF scales. These scales jointly explained 31% of the variance in Preferred Yellow. The majority of the variance was explained by the Group-Oriented scale (β = .72). The Warmth (β = .49) and Gregariousness (β = .35) scales explained further meaningful variance in Preferred Yellow.

Social Yellow demonstrated a Multiple R of .47 with the 16PF Warmth and Friendliness scales, jointly explaining 22% of the variance in Social Yellow. The majority of the variance was explained by the Warmth scale (β = .71).

These findings suggest that these 16PF scales contribute meaningful variance to the Preferred Yellow and Social Yellow scales as hypothesized.

Blue Scales with Multiple Hypotheses

The Preferred Blue scale demonstrated a Multiple R of .50 with the 16PF Sensitivity and Complexity scales, jointly explaining 25% of the variance in Preferred Blue. A majority of the variance explained was due to the Complexity scale (β = .92).

Social Blue demonstrated an adjusted Multiple R of .52 with the 16PF Intellect and Complexity scales. These scales jointly explained 27% of the variance in Social Blue scale. A majority of the variance was explained by the Complexity scale (β = .89) while Intellect also explained some variance in the Social Blue scale (β = .21).

Instinctive Blue demonstrated an adjusted Multiple R of .54 with the Self-assured, Dutifulness, Warmth, Intellect, and Emotional Stability 16PF scales, jointly explaining 29% of the variance in Instinctive Blue. Of the five scales, Self-assured (β = -.68) explained the most variance in the Instinctive Blue scale, while Dutifulness (β = .35) and Warmth (β = .33) also contributed meaningful variance.

These findings suggests that the hypothesized relationships contribute meaningful variance to all three Blue scales.

Green Scales with Multiple Hypotheses

Social Green had an adjusted Multiple R of .43 with the Dutifulness and Group-Oriented scales. These scales jointly explained 20% of the variance in Social Green. Both the Dutifulness (β = .72) and Group-Oriented (β = -.74) scales explained meaningful variance in the Social Green scale. Dutifulness was positively related, and Group-Oriented was inversely related to Social Green, as predicted. The inverse relationship suggests this scale may tap into Self-Reliance, which is the opposite pole of Group-Oriented in the 16PF model.

Instinctive Green had an adjusted Multiple R of .21 with the Orderliness, Imagination, Group-Oriented, and Gregariousness 16PF scales, jointly explaining 4% of the variance in Instinctive Green. The Orderliness (β = .19) and Imagination scales (β = -.18) explained a similar amount of variance in Instinctive Green. Orderliness was positively related, and Imagination was inversely related to Instinctive Green.

The findings suggest that the hypothesized relationships contribute substantial variance to the Social Green scale, and a smaller but meaningful amount of variance to the Instinctive Green scale.

Table 3
Linear Regression Analysis of Hypothesized Relationships Between Personalysis and 16PF Scales

Yellow	Scales	β	Multiple R	Variance explained
Rational	Warmth	.49	.56	31%
	Gregariousness	.35		
	Friendliness	-.07		
	Forthright	-.03		
	Group-oriented	.72		
Social	Warmth	.71	.47	22%
	Friendliness	.21		
Blue				
Rational	Sensitivity	.02	.50	25%
	Complexity	.92		
Social	Intellect	.21	.52	27%
	Complexity	.89		
Instinctive	Warmth	.33	.54	29%
	Intellect	-.30		
	Emotional Stability	-.19		
	Dutifulness	.35		
	Self-assured	-.68		
Green				
Social	Dutifulness	.72	.45	20%
	Group-oriented	-.74		
Instinctive	Gregariousness	.01	.21	4%
	Imagination	-.18		
	Group-oriented	-.14		
	Orderliness	.19		

Notes. N = 423. β denotes unstandardized regression coefficient.

Multitrait-Multimethod Matrix Analysis

A multitrait-multimethod matrix was computed to further evaluate the construct validity of the Personalysis assessment with the 16PF. In Table 4, there are two statistics used to demonstrate convergent and discriminant validity organized by Personalysis scale. The first statistic, hypothesized, is the largest correlation observed between each Personalysis scale and the theoretically related 16PF scale(s). When more than one hypothesis was made for a scale, the largest univariate correlation was used as this relationship accounts for the most variance in that Personalysis scale. When only one hypothesis was made for a scale, that univariate correlation was entered into the multitrait-monomethod matrix. Second, the average of the non-hypothesized correlations is reported to demonstrate discriminant validity with theoretically unrelated scales. Evidence of convergent and discriminant validity exists when hypothesized correlations are *higher* than non-hypothesized scales.

Red Scales

For the Preferred Red scale, the hypothesized correlation of .30 with Assertiveness was higher than the non-hypothesized average correlation ($\bar{r}$ = -.05). These findings provide strong support for the Preferred Red scale.

Social Red demonstrated a -.03 hypothesized relationship with Assertiveness, which is smaller and lower than the -.13 non-hypothesized average correlation. Thus, no support is found for the Social Red scale.

The Instinctive Red scale demonstrated a strong hypothesized relationship ($\bar{r}$ = .42) with the Assertiveness scale, which is higher than the non-hypothesized average correlation ($\bar{r}$ = .09). Thus, strong support is found for the Instinctive Red scale.

In sum, strong evidence of construct validity was found for the Preferred and Instinctive Red scales, and no support for the Social Red scale.

Yellow Scales

Preferred Yellow had a .50 correlation with the Group-Oriented scale, which is higher than the non-hypothesized average correlation ($\bar{r}$ = .08). Strong construct validity evidence was found for the Preferred Yellow scale.

Social Yellow had a .46 correlation with the Warmth scale. This correlation is higher than the .17 non-hypothesized average correlation observed, providing strong construct validity evidence for the Social Yellow scale.

The Instinctive Yellow scale demonstrated a .10 correlation with the Group-Oriented scale which does not meet the threshold for meaningfulness, but is higher than the non-hypothesized correlation of .00. Thus, there was mixed support for the Instinctive Yellow scale.

In sum, strong evidence of construct validity was found for the Preferred and Social Yellow scales, and mixed support for the Instinctive Yellow scale.

Blue Scales

Preferred Blue demonstrated a .50 correlation with the Complexity scale. This correlation is higher than the .08 non-hypothesized average correlation, thus providing strong support for this scale.

Social Blue had a .52 correlation with the Complexity scale, which is higher than the average non-hypothesized average correlation of .05. Thus, strong construct validity evidence is found for the Social Blue scale.

Lastly, the Instinctive Blue scale had a .48 correlation with the Self-Assured scale which is higher than the -.11 correlation with non-hypothesized scales, providing strong evidence in support of the Instinctive Blue scale.

In sum, convergent validity evidence was strong for the Preferred, Social, and Instinctive Blue scales.

Green Scales

The Preferred Green scale demonstrated a marginal .19 correlation with the Orderliness scale. The non-hypothesized correlation was somewhat larger in magnitude but negative in direction (r = -.27), suggesting a negative relationship with several non-hypothesized scales (see Table 5). Thus, some validity evidence is found for the Preferred Green scale with the 16PF when considering exploratory scales.

Social Green demonstrated a .31 correlation with the Dutifulness scale. This correlation is higher than the -.24 non-hypothesized average correlation, providing evidence of construct validity for the Social Green scale.

Lastly, the Instinctive Green scale demonstrated a .18 correlation with the Orderliness scale. This correlation is substantially above the non-hypothesized correlation of .06, providing support for the Instinctive Green scale.

In sum, evidence of construct validity is observed for the Social and Instinctive Green scales, and some evidence is observed for the Preferred Green.

Table 4
Personalysis Scale Correlations With 16PF Scales

Red	Hypothesized*	Non-hypothesized
Preferred	.30	-.05
Social	-.03	-.13
Instinctive	.42	.09
Yellow		
Preferred	.50	.08
Social	.46	.17
Instinctive	.10	.00
Blue		
Preferred	.50	.08
Social	.52	.05
Instinctive	-.48	-.11
Green		
Preferred	.19	-.27
Social	.31	-.24
Instinctive	.18	.06

Notes. *N* = 423. * Denotes largest hypothesized correlation.

Exploratory Findings

In the exploratory phase, large non-hypothesized correlations were examined between Personalysis and 16PF scales. Because these analyses are exploratory, a higher threshold of $r \geq .30$ was used. As shown in Table 5, post hoc analyses revealed that the 16PF Assertiveness scale demonstrated meaningful correlations with Social Blue (r = .33) and Instinctive Blue (r = -.45). The 16PF Complexity scale was correlated with Social Blue (r = -.34). Further, 16PF Gregariousness was correlated with Social Yellow (r = .35) and Green (r = -.33). The 16PF Group-oriented scale was correlated with Social Yellow (r = .54) and Preferred

Green (r = -.31). The 16PF Relaxed scale was correlated with Preferred (r = -.30) and Social Green (r = -.32). Further, 16PF Self-assured was correlated with Instinctive Red (r = .31). Moreover, 16PF Trusting was correlated with Preferred Yellow (r = .31).

Table 5
Exploratory Correlations Larger Than .29 Between Personalysis and 16PF Scales

Red	Scales	Correlation
Instinctive	Self-assured	.31
Yellow		
Preferred	Trusting	.31
	Gregariousness	.35
Social	Forthright	.33
	Group-oriented	.54
Blue		
Social	Assertiveness	.33
	Complexity	.34
Instinctive	Assertiveness	-.45
Green		
Preferred	Group-oriented	-.31
	Relaxed	-.30
Social	Gregariousness	-.33
	Relaxed	-.32

Note. *N* = 423.

Discussion

The construct validity of the Personalysis scales was investigated using several methods. First, each scale was evaluated for convergent validity by examining their relationships to theoretically linked 16PF scales. 26 a priori hypotheses were developed across all 12 Personalysis scales. 19 of the 26 relationships observed, or 63%, of the hypotheses were supported by univariate correlations using an established threshold of .20. Further, 10 of the 12

Personalysis scales tested, or 73%, were found to be supported by a correlation of .18 or higher with at least one hypothesized 16PF scale.

When more than one 16PF scale was hypothesized to relate to a single Personalysis scale, a series of linear regressions revealed that the median amount of variance jointly explained by the theoretically predicted scales was 25% (range 4% to 31%). Substantial variance ($\geq$ 20%) was explained for six of the seven multivariate predictions. The 16PF explained a smaller amount of the variance in the Instinctive Green scale.

A multitrait-monomethod matrix showed that 9 of the 12 (75%) a priori hypotheses were further supported by substantially higher correlations with hypothesized than non-hypothesized 16PF scales. Social Red and Preferred Green scales showed lower hypothesized than non-hypothesized correlations. Post hoc exploratory analyses revealed that elements of the 16PF Group-oriented and Relaxed are negatively related to the Preferred Green scale. Further, the 16PF scales Gregariousness, Forthright, and Group-oriented were related to Social Red. Although unanticipated, these post hoc findings are consistent with Personalysis theory, and serve to better articulate it's nomological network. Furthermore, Instinctive Yellow demonstrated a notable trend toward convergent validity with the group-oriented scale (r =.10), but the magnitude of the finding did not exceed our established threshold.

In conclusion, the present study found strong convergent validity for 9 of the 12 Personalysis scales. Although convergent validity was unsupported for Social Red, convergent validity evidence was found with the MBTI. Further, some support for the Preferred Green scale was found with the 16PF and MBTI. Further research may explore which scales from previously established personality assessments are most related to Instinctive Yellow and Preferred Green.

Evaluation of a test's validity is an ongoing process involving a collection of different lines of evidence across multiple studies. The results of this study provide strong support for a majority of the Personalysis scales, and some support for others. The results presented here should be interpreted in context with results of the previously reported validity study with MBTI (see Table 6 in Appendix). Taken together, these two studies provide some convergent validity evidence for 11 of the 12 Personalysis scales.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Gatewood, R. D., Field, H. S., & Barrick, M. (2013). *Human Resource Selection* (7th Ed.). Cengage Learning.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.

Haldun, A. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(3), 91–93. https://doi:10.1016/j.tjem.2018.08.001

Myers, I. B. & Briggs, K. C. (1998). *The Myers-Briggs Type Indicator Form M*. Consulting Psychologist Press.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sackett, P. R., Tippins, N. T., Arthur, W., Jr, Delany, T., Dunleavy, E. M., Hayes, T. L., Hough, L., Oswald, F., Putka, D. J., Ryan, A. M., & Schmitt, N. (2018). *Principles for the Validation and Use of Personnel Selection Procedures*. American Psychological Association.

Smith, G. T. (2005). On Construct Validity: Issues of Method and Measurement. *Psychological Assessment, 17*(4), 396–408. https://doi.org/10.1037/1040-3590.17.4.396

Appendix

Table 6

Convergent Validity Support for each Personalysis Scale

Red	MBTI	16PF
Preferred	Unsupported	Supported
Social	Supported	Unsupported
Instinctive	Supported	Supported
Yellow		
Preferred	Supported	Supported
Social	Supported	Supported
Instinctive	–	Some support
Blue		
Preferred	Supported	Supported
Social	Supported	Supported
Instinctive	Supported	Supported
Green		
Preferred	Some support	Some support
Social	Supported	Supported
Instinctive	Supported	Supported

Notes. MBTI *N* = 295. 16PF *N* = 423.