PERSONALYSIS®

Personalysis Technical Report #2: Methodology of Personalysis Studies

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 1st, 2023

Personalysis Corporation®, 675 Bering Drive, Houston, TX 77057. USA. www.personalysis.com

Contents

Introduction

The purpose of the Personalysis research study is to evaluate the basic psychometric properties of the Personalysis assessment instrument. Four analyses were conducted to evaluate: internal consistency, stability of scores over time, and construct validity as it relates to two other commonly used personality tests. The data collection methods were the same for all four analyses. To avoid repetition, this technical report describes the design, data collection procedures, study demographics, and measures for the entire study. Data analyses, results and conclusions are described separately for each study in technical reports 3 – 5. Complete demographic percentages for each of the construct validity studies may be found in the appendix of this technical report.

Design

The purpose of this research study is to measure the construct validity and reliability of the 12 Personalysis scales. When measuring the construct validity and reliability of an assessment it is imperative to establish the external validity of the research studies. External validity is defined as the generalizability of study results to different settings, samples, and times (Steckler & McLeroy, 2008). External validity is important when conducting research to generalize the results to the population of interest (Campbell, 1957). In simpler terms, the results of these studies are intended to generalize to Personalysis clients.

Temporal reliability was evaluated using a within-person approach, where Personalysis scores are compared across two time points. Two established measures of personality, namely the Myers-Briggs Personality Inventory (MBTI) and the Sixteen Personality Factor Questionnaire (16PF), were used to establish the construct validity of the Personalysis scales.

A viable crowdsourcing platform was needed to collect participant data. When comparing crowdsourcing platforms, the following qualities were considered: a) representativeness of the Personalysis client population; b) convenience to collect data; c) ability to collect data on a large and diverse group; d) cost effectiveness. MTurk was chosen as the crowdsourcing platform given its track record of providing high quality data (Hauser et al., 2019; Zhang & Gearhart, 2020), diversity in sampling (Smith et al., 2015), and efficiency/cost effectiveness (Antoun et al., 2015). In addition to MTurk, CloudResearch was used to filter non-purposeful respondents to ensure data quality.

A two-timed study methodology was conducted to evaluate the reliability of the Personalysis scales. Specifically, the Personalysis assessment was administered at time one, then two weeks later the Personalysis assessment was again administered to participants. The purpose of this design is to measure how consistent participant's responses are across a two–week period.

Two reliability metrics were used to assess each Personalysis scale's reliability. First, test-retest reliability was analyzed by correlating participant's time one scale scores with their time two scale scores. Second, a classification consistency analysis was performed by comparing the Personalysis scale interpretations across two time points. The traditional Personalysis scoring system provides the user a color designation reflecting their dominant trait for individuals

who score a 3.0 or higher on a scale. These color designations are the primary source for interpreting test results and providing feedback to test takers during workshops or individual debriefs conducted by certified Personalysis practitioners. To evaluate classification consistency, participants were labeled as a color using the traditional scoring system described in the measures section (below) on Personalysis. For example, if a participant scores a 3.0 or higher on the Social Red scale, they will be classified as Social Red. If an individual scores a 2.5 or lower on the Social Red scale, they will *NOT* be classified as Social Red. Thus, classification consistency evaluates whether participants receive the same color (i.e., trait) interpretation across two test intervals.

As discussed above, the project team selected two widely used personality assessments to assess convergent and discriminant validity of the Personalysis scales. All participants received the Personalysis assessment first, then participants were randomly split into groups either receiving the MBTI or 16PF assessment.

Procedure

An initial demographic screener survey was used to verify the demographic composition of the study sample. Across both demographic screeners, a total of 1,408 participants signed up for the study and filled out a demographic questionnaire. Once demographic information was obtained, participants were invited to take one of two comparative surveys containing the Personalysis assessment with either the MBTI or 16PF. A total of 425 Participants were invited to participate in the Personalysis–MBTI comparative survey. Of those 425 participants, 295 (69%) of participants responded to the invitation, passed the attention checks, and took both the time 1 and time 2 Personalysis surveys and the MBTI survey. A total of 525 Participants were invited to take the Personalysis–16PF comparative survey. Of those 525 participants, 423 (81%) of participants responded to the invitation, passed the attention checks, and took both the time one and time two Personalysis surveys and the 16PF survey. Across both comparative surveys, the average completion/response rate was 76%. A flow chart of the procedure can be found in Figure 1 below.

Participants were invited to take either the Personalysis–MBTI or Personalysis–16PF comparative survey, they were not permitted to take both. There was no duplication in the sample for either comparative survey. The Personalysis assessment was always administered first, while the MBTI or 16PF was administered second.

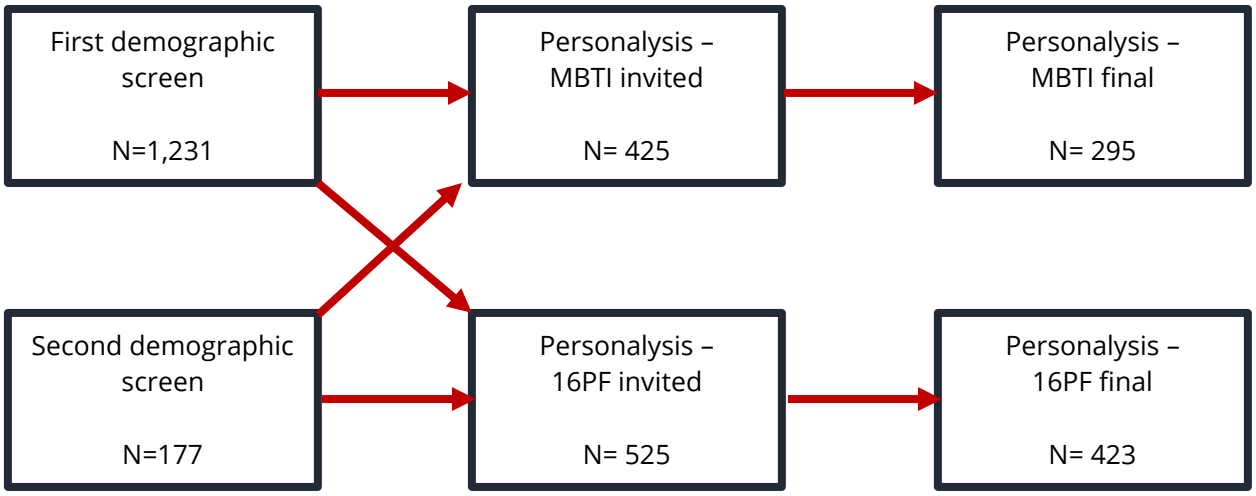
As discussed above, MTurk was used as the primary crowdsourcing platform. Participants were compensated $5 for completing the first comparative survey. If participants completed the initial comparative survey while passing over two-thirds of the attention checks, they were invited to take the Personalysis assessment a second time. Participants completed the second assessment an average of 13 days after the first assessment. This second survey did not contain the MBTI nor 16PF, and therefore took less time for participants to complete. Participants were compensated an additional $5 for completing the second Personalysis survey and passing over two-thirds of the attention checks.

The three authors of this report and an expert in Personality research were involved in the development of the a-priori hypotheses for the evaluation of construct validity. Each of the members have extensive experience with either one of or multiple of the following personality assessments: Personalysis, 16PF, MBTI instruments. These four members totaled over 58 years of experience researching and using personality assessments. Before hypothesis development, each author reviewed the items of the three Personality assessments. After reviewing the items, the four-person team collaboratively developed hypotheses for each Personalysis scale.

During hypothesis development a question was raised regarding what constitutes a considerable relationship between two scales. Initially, statistical significance was suggested as the standard to determine whether a Personalysis scale correlating with another personality scale demonstrated convergent validity. As any relationship can be statistically significant with large enough of a sample size (Lykken, 1968), a decision to avoid statistical significance was made. Rather, an a-priori correlation threshold of .20 was established as the criterion to confirm hypotheses that two scales are meaningfully related. This follows the aim of the study to establish whether Personalysis scales demonstrate a meaningful relationship with other related personality scales (Kirk, 1996).

Figure 1

Example of Procedure



Participants

Data Quality Control

As described above, participants were recruited from MTurk, an online crowdsourcing data collection platform. MTurk has the benefit of a large group to sample from, diverse participant pool, ease of use, quick data collection, and flexibility for research design (Aguinis et al., 2021).

In addition to MTurk acting as the primary tool for crowdsourcing, CloudResearch is a partner of the MTurk platform that filters data quality of participants. Previous findings have shown that when using CloudResearch's data quality filter, study results are superior to that of Prolific and other popular crowdsourcing platforms (Hauser et al., 2022; Litman et al., 2021). Thus, the CloudResearch platform was used in the current study to improve the internal validity of study findings.

Several data quality measures were implemented to collect data from participants with a history of high-quality data. First, participants included in the study were vetted through CloudResearch's data quality protocol. Second, participants must have completed at least 100 tasks (HITs) on MTurk with an approval rating of 95% or higher to be allowed to take the survey. If a participant met these two requirements, they were able to participate in this study. This led to 1,231 MTurkers, from now on called participants, who signed up to participate in the study. After collecting the first round of data, the demographic breakdown did not meet the demographic target set before the study was conducted. To address this lack of diversity in participant recruiting, an additional 177 participants who reported as either Black/African-American, Hispanic/Latino, or Asian/Asian-American were recruited through MTurk and invited to participate through the second demographic screener.

There was a three-wave approach to data collection. At wave one, participants took a demographic screener consisting of race, ethnicity, gender, age, and education level to verify demographic information. A common concern in convenience sampling is the chance participants may lie about their demographics to increase their likelihood of being selected to the next phase of the study (Smith et al., 2015). To address this concern, participants were told they would "be selected based on the demographic breakdown" of the sample, and not informed

of the study's target demographics. This message was used to incentivize participants to respond honestly about their demographics.

Participants were included in the final analysis if they took both the time one and time two assessments. A total of 385 participants took the Personalysis–MBTI. 90 participants (23%) were removed from the original Personalysis–MBTI group. Further, a total of 548 participants took the Personalysis–16PF survey. 125 participants (23%) were removed from data analysis for reasons discussed below.

Participants were removed from data analysis for any of the following four reasons: 1) failing attention checks; 2) finished survey in less than 10 minutes; 3) did not report their MTurk ID; 4) randomly removed to reach demographic targets.

Attention checks

First, participants were removed if they were unable to pass 67% or more of the attention checks. There were two types of attention checks: i) items that were obviously correct (e.g., Select "nice" below) and; ii) two identical items located within a few questions of one another. The purpose of the first attention check is to identify individuals who did not read each item. The second type of attention check identifies if participants considered each item carefully. If a participant does not endorse the same item just a few moments later, it is likely that they were not paying sufficient attention. If an individual did not pass the at least two-thirds attention checks, their data were discarded, and they were removed from the invitation list at time two.

Survey time

Second, participants were removed if they completed the survey in an unreasonably short amount of time. A threshold of 10 minutes was set as the minimum expectation to complete the assessment given the amount of time each participant would need to read and think about their answer to the question.

Unique Identification

Third, participants who did not provide their MTurk ID were removed from the analysis.

Demographic mix

Lastly, some participants were randomly removed if a particular demographic category had been oversampled in relation to U.S. Census

demographic estimates. These participants were paid in full at time one but were not invited at time two.

After all of the attention checks and inclusion criteria, a total 718 participants remained for data analysis. Table 1 shows the demographic percentages of the entire sample. The U.S. Census was chosen to align with the Personalysis client target, which is working adults in the United States. The U.S. Census estimates below reflect the working population (U.S. Bureau of Labor Statistics, 2021) which differs slightly from the general population of all U.S. citizens. All census estimates were filtered to include employed citizens with some college or higher, consistent with the current client base for Personalysis team building and coaching workshops.

Below are the demographics of the entire 718 participant database. The reliability analyses use all 718 participants. Separate demographics for the MBTI and 16PF comparative surveys can be found in the appendices under Tables 5 and 6, respectively. These demographics should be considered when evaluating the results of both construct validity reports [insert report links here].

Study Demographics

Gender

Gender findings indicate that the sample is within 1% of census estimates. Males represented 52% of the sample, while Females represented 46% of the sample. A total of 16 participants (2%) identified as non-binary/non-conforming (*n* = 12), or transgender (*n* = 4). The census does not currently report all gender categories, instead they report sex as a binary statistic. In this study female and male census estimates were used for the female and male targets even though our survey allowed for other gender options.

Ethnicity

Ethnicity estimates align exactly with the census estimate for working adults. A total of 83 (12%) participants reported as Hispanic or Latino, while 634 (88%) of participants reported they were not Hispanic nor Latino.

Race

There were a total of six options available to participants regarding race. Participants could indicate whether they were American Indian or Alaska Native,

another race, Asian or Asian American, Black or African American, Hispanic or Latino, two or more races, and White or Caucasian. The U.S. Census does not include American Indian/Alaska Native, two or more races, or another race as options. Although the target demographics use the U.S. Census estimates, the purpose of each study is to apply the results to all of Personalysis clients. Therefore, three additional categories were included to generalize to the broader Personalysis population. Further, providing additional options allows participants to describe their race as specifically as possible.

The findings indicate that two participants (<1%) reported as American Indian or Alaska Native. Three participants (< 1%) reported they were another race. A total of 58 participants (9%) reported as Asian or Asian American. This estimate is within one percent of the reported demographics of the census estimate for working adults which, again, may differ from the general population. 78 participants (12%) reported themselves as Black or African American, which is the same percentage as the census estimate. 27 participants (4%) reported as two or more races, although there is no census estimate of two or more races. Finally, a majority (74%) of the participants reported themselves as White or Caucasian. Lastly, this number is 3% lower than the census estimate reflecting the diversity of the current sample.

Age

Reported age was split into four groups to match the census estimates below. A total of 42 participants (6%) were in the 20-24 age group. This number is 3% lower than the census estimate. A total of 554 (77%) of participants were in the 25-54 age group. This number is higher than the census estimate by 11%. 88 participants (12%) fell into the 55-64 age group, which is 6% lower than the census estimate. Lastly, 15 participants (5%) were in the over 65 age group. This finding is 2% lower than the census estimate. Overall, the finding that this MTurk sample is 11% higher than the 25-54 age group census estimate is in line with previous research finding that MTurkers tend to be younger than the national average (Ipeirotis, 2012). Given that the age of Personalysis clients tends to be slightly younger, the demographic findings presented here are largely consistent with expectations for working adults.

Education Level

There were a total of nine responses options participants could select to indicate their highest level of education achieved. The first two options were less

than high school or other type of education specified where the participant could write their highest education level received. These two options were combined to form the less than high school or other category. The second category includes individuals who reported their highest education level as high school graduate. Participants could select from three response options of some college, post vocational school, or associate's degree which were combined to form a composite category.

Lastly, participants could select bachelor's, master, or doctorate degrees. These three options were combined to form a bachelor's or higher category.

Four participants (1%) reported they had less than a high school degree or other educational level. This is 7% lower than the census estimates. 44 participants (6%) reported they have a high school degree, which is 17% lower than the census estimates. There were 250 participants (35%) who reported they completed some college, an associate's degree, or completed post high school vocational training. This estimate is 6% higher than the census estimates. Lastly, 420 participants (58%) reported they earned a bachelor's degree or higher, which is 17% higher than the census estimates.

Overall, the targeted sample was intended to reflect a majority of participants with a bachelor's degree. Personalysis clients tend to be working individuals with higher education levels taking the Personalysis assessment. Thus, this targeted sampled generalizes well to the Personalysis client population.

Table 1

Demographics of MTurk Sample

Gender	*n*	%	Census estimates
Female	332	46%	47%
Male	370	52%	53%
Other*	16	2%	n/a
Ethnicity			
Hispanic	83	12%	12%
Non-Hispanic	634	88%	88%
Race			
American Indian or Alaska Native	2	<1%	n/a
Another race	3	<1%	n/a
Asian or Asian American	58	9%	8%
Black or African American	78	12%	12%
Two or more races	27	4%	n/a
White or Caucasian	476	74%	77%
Age			
20-24	42	6%	9%
25-54	554	77%	66%
55-64	88	12%	18%
over 65	34	5%	7%
Educational Attainment			
Less than high school or other	4	1%	8%
High school graduates	44	6%	23%
Some college, associate's degree, or vocational training	250	35%	29%
Bachelor's degree or higher	420	58%	41%

Notes. *Other includes non-binary/non-conforming, transgender man, transgender woman, and prefer not to answer. *N* = 718.

Measures

A total of three personality assessments were analyzed: Personalysis, the Meyers-Briggs Type Indicator (MBTI), and the Sixteen Personality Factor Questionnaire (16PF).

Personalysis

The Personalysis (Noland, 1975; Noland, 2005) assessment is an ipsative, forced-choice personality measure that presents two statements (adjectives or phrases) per item. Each item is designed to allow the respondent to choose which of two statements is most like them. There are a total of 12 scales in the Personalysis assessment. Four overarching personality traits are measured across three contexts (commonly referred to as Dimensions). To avoid pejorative labels, the four traits are designated by the colors Red, Yellow, Blue and Green (commonly referred to as Colors). The three Dimensions are: Preferred, Social, and Instinctive.

The Personalysis assessment was developed in the mid 1970's by James R. Noland, a behavioral scientist who specialized in management and organizational behavior. Personalysis integrates and updates the concepts of Berne, Freud, Jung, Maslow, W.I. Thomas and MacLean. The Color designations are based on Jungian dichotomies of Extroversion(E)–Introversion(I), Sensing(S)–Intuition(N), Thinking(T)–Feeling(F), and Judging(J)–Perceiving(P). The work of W.I. Thomas' four fundamental desires of mastery, recognition, new experiences/stimulation, and security enhanced the definitions of colors. The Dimensions are based on Freud's id, ego, and superego that were further developed by Berne into child, parent, adult. Maslow is reflected in the design as each dimension addresses a variable in his Hierarchy of Needs.

The Personalysis items are focused on contextualized work-related situations. One unique aspect of the Personalysis assessment is the measurement of the three Dimensions for each of the four color-coded traits. Further, scales with the same color designation (e.g., Preferred Red, Social Red, and Instinctive Red) are all related to one another, but the specific trait being measured is considered within a different dimensional context. The Preferred dimension measures the type of work, and approach to work, an individual enjoys and finds the most rewarding. Social dimension measures an individual's expectations of appropriate behavior and communication in social situations. And the Instinctive dimension measures what builds an individual's confidence when making decisions, responding to change or in times of extreme stress.

Contextualized personality assessments have been shown to demonstrate higher predictive (Swift & Peterson, 2019) and construct validity (Golubovich et al., 2020) than non-contextualized assessments. Furthermore, the contextual items in the Personalysis scales should lead to clear relationships between the item endorsed and the behavior at work. For more information regarding the history of Personalysis and the theories upon which it was developed, please see Personalysis Technical Report #1: History of Personalysis.

Table 2

Example of Personalysis Scales

	Description	*Example item*
	Red	
Preferred	Initiates immediate action Enjoys a fast pace Possesses high energy Practical Enjoys challenges Focused on rapid accomplishment	active-practical
Social	Straightforward Progress focused Provides direction Assertive	be direct, straightforward
Instinctive	Sets clear goals Relies on personal experience Activity-oriented Desires freedom to take action Has a high sense of urgency	intense; fiery
	Yellow	
Preferred	Helps and supports others Collaborative People-oriented Facilitates outcomes Influences and inspires	receptive, collaborative
Social	Outgoing, expressive Invite others into the conversation Connects to others on a personal level Accommodating Asks for others help and opinions	encourage cooperation
Instinctive	Opportunistic Likes to know who is involved Gathers opinions, discusses possibilities Flexible and influential	too permissive, gullible

Table 2

Example of Personalysis Scales

	Description	*Example item*
Blue		
Preferred	Utilizes experimental solutions Explores information/ideas Diagnoses complex issues Deep learner Innovative	thoughtful; curious
Social	Expressive Asks for input Shares context, explains Ask questions Inquisitive	be informed; ask questions
Instinctive	Needs context and rationale Thinks and reflects Holds multiple views, perspectives Clarifies and contemplates alternatives	theoretical; impractical
Green		
Preferred	Focused on structure and order Relies on systems Plans and prioritizes Follows proven processes Creates stability	systematic; orderly
Social	Formal, literal Provides structure Process focused Specific, exact Schedules and coordinates	be responsible; maintain stability
Instinctive	Predictable Detail-oriented Identifies specific responsibilities Analyzes and reduces risks	critical; look for errors

MBTI

The Myers-Briggs Type Indicator (MBTI®) Form M was used to assess construct validity in Personalysis Technical Report #4. Similar to the Personalysis assessment, the Myers-Briggs Type Indicator is a forced-choice, ipsative measure that forces respondents to choose between two options.

The MBTI has four bi-polar scales in which the test taker is forced to choose between two opposite ends of a spectrum. These four bipolar scales are Extroversion(E)–Introversion(I), Sensing(S)–Intuition(N), Thinking(T)–Feeling(F), and Judging(J)–Perceiving(P). Participants receive a type classification based on the 16 types possible. For example, if a respondent scored higher on the Extroversion (E), Intuition (N), Feeling (F), and Judging (J) scales, they would be classified as ENFJ.

A major difference between the MBTI and Personalysis assessment is how many scales are scored in each item. In the MBTI, each response option is unidimensional, meaning only one scale is measured per item. Conversely, the Personalysis items are multidimensional, meaning statements from different scales comprise the two opposing options for each item.

Alpha scale reliabilities for each of the four MBTI scales are .97 in the current sample. In addition to its similarity to the Personalysis assessment, the MBTI has demonstrated strong construct validity (Thompson & Borello, 1986), and is one of the most widely used personality assessments (Davito, 1995; Furnham, 1996).

Table 3

Descriptions and Example Items of MBTI Scales

	Description	Example Statement
Extroversion	exerts energy toward those around them stimulated by the environment open to new experiences action-oriented sociable	Open
Introversion	thinks through ideas before discussing them focused on the internal, subjective state interested in concepts and ideas enjoys privacy	Private
Sensing	remembers past and present details well focused on the immediate moment enjoys the present moment observant	Facts
Intuition	future-oriented imaginative theoretical abstract creative	Ideas
Thinking	focused on facts in decision making analytical logical objective	Objective
Feeling	considers others when a decision is made uses their feelings in decision making concerned about others subjective	Warm
Judging	concerned with making decisions organizes activities plans operations seeks closure	Orderly
Perceiving	hesitant to make an immediate judgment desires as much information as possible receptive to new information spontaneous curious	Easygoing

Note. Descriptions and example statements taken from the MBTI® Form M Manual.

Sixteen Personality Factor Questionnaire (16PF)

The 16PF was used to assess construct validity in Personalysis Technical Report #5. Similar to the MBTI, the 16PF has been widely used around the world (Butcher & Rouse, 1996). Creation of the 16PF assessment in the early 1900s took a different methodological approach than other personality assessments of its time. Raymond Cattell, the founder of the 16PF, used self-report, peer ratings, and objective measures in addition to a statistical technique called factor analysis to develop the 16PF questionnaire (Cattell & Schuerger, 2003). Cattell found there were 16 primary traits representing personality within his studies (Cattell & Cattell, 1995). Later studies showed that the 16 scales could be organized into five global dimensions (Cattell & Mead, 2008), consistent with the Five-Factor Model (FFM). The current study uses the original 16PF scales which are commonly used in practice.

In the current study, the 16PF was derived from a 160-item International Personality Item Pool (IPIP) representation of the 16PF (Conn & Rieke, 1994). These 16PF replication scales utilize a Likert-type response scale to measure each respondent's answer. Response options ranged from 1 "Strongly disagree" to 5 "Strongly agree." Alpha reliabilities of the 16PF scales ranged from .72 to .94, with an average alpha reliability of .85 in the current sample.

Table 4
Descriptions, Example Items, and Alpha Reliabilities of 16PF Scales

	Factor	High description	Low description	Example Item	Alpha Reliability
Warmth	A	warm outgoing kind	cool reserved impersonal	Know how to comfort others.	.89
Intellect	B	abstract thinking bright quick to grasp ideas	concrete thinking dull slow to grasp ideas	Learn quickly.	.75
Emotional Stability	C	emotionally mature mature calm	affected by feelings easily annoyed emotional	Feel comfortable with myself.	.93
Assertiveness	E	dominant assertive competitive	submissive humble easily led	Take control of things.	.89
Gregariousness	F	enthusiastic expressive cheerful	restrained introspective serious	Love large parties.	.81
Dutifulness	G	conscientious rule-bound responsible	expedient casual disregards rules	Try to follow the rules.	.83
Friendliness	H	bold sociable spontaneous	shy timid hesitant	Make friends easily.	.94
Sensitivity	I	tender-minded sensitive refined	tough-minded rough realistic	Enjoy discussing movies and books with others.	.72
Trusting	L	easy to get along with cheerful concerned about others	suspicious distrustful skeptical	Trust what people say.	.93
Imagination	M	creative absorbed in thought unconventional	practical unimaginative steady	Like to get lost in thought.	.81

	Factor	High description	Low description	Example Item	Alpha Reliability
Forthright	N	open genuine sentimental	socially aware calculated unsentimental	Show my feelings.	.92
Self-assured	O	secure mature confident	insecure constantly worrying anxious	Am not easily bothered by things.	.84
Complexity	Q1	open to change well informed interested in new ideas	resistant to change traditional cautious of new ideas	Enjoy hearing new ideas.	.84
Group-oriented	Q2	listens to others goes along with the group seeks approval from others	self-sufficient prefers own decisions resourceful	Enjoy teamwork.	.76
Orderliness	Q3	disciplined strong control of emotions prefer order	careless of social rules unbothered by social demands impulsive	Like order.	.85
Relaxed	Q4	tranquil composed satisfied	tense impatient frustrated	Am not easily annoyed.	.88

Notes. Descriptions taken from the 16PF Manual (Form E). Items taken from https://ipip.ori.org.

Discussion

The purpose of this research study is to measure the construct validity and reliability of the 12 Personalysis scales. Furthermore, external validity needed to be established to ensure the current study generalizes to different settings, samples, and times. Using an MTurk sample to collect data on employed U.S. citizens who have at least some college education represents a target sample within reasonable estimates of Personalysis clients. Census estimates matching the Personalysis client population were evaluated, and current study demographics are within the target population of the study. Therefore, the external validity of the current study generalizes well to the Personalysis client population.

In addition to generalizing to current Personalysis clients, an effort was made to target a future audience. Sample estimates of race, gender, ethnicity, and age all fall within the census estimates. Furthermore, some individuals who completed less than high school, high school, post vocational school, or other type of education were included to target a wider sample that may apply to future Personalysis clients.

Further, careful quality controls techniques (described fully above) were implemented with data collection procedures, allowing for confidence in the integrity of the dataset, and the accuracy of subsequent findings.

Overall, the demographics of the current sample align well with the Personalysis client base which is comprised of educated working adults. The characteristics of the study sample and the quality of the data collection procedures allow for generalizability of the specific findings reported in each of the subsequent technical reports 3-5.

References

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management, 47*(4), 823–837. https://doi.org/10.1177/ 0149206320969787

Antoun, C., Zhang, C., Conrad, F. G., & Schober, M. F. (2016). Comparisons of online recruitment strategies for convenience samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk. *Field Methods, 28*(3), 231–246. https://doi.org/10.1177/1525822X1560314

Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology, 47*, 87–111. https://doi.org/10.1146/annurev.psych.47.1.87

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*(4), 297-312. https://doi.org/10.1037/h0040950

Cattell, H. E., & Mead, A.D. (2008). The Sixteen Personality Factor Questionnaire (16PF). In G. Matthews, D. H. Saklofske, & G. J. Boyle (Eds.), *The SAGE Handbook of Personality Theory and Assessment*. Sage Publications.

Cattell, H. E., & Schuerger, J.M. (2003). *Essentials of 16PF Assessment*. Sage Publications.

Cattell, R. B., & Cattell, H. P. (1995). Personality structure and the new fifth edition of the 16PF. *Educational and Psychological Measurement, 55*(6), 926–937. https://doi.org/10.1177/0013164495055006000

Conn, S. R., & Rieke, M. L. (1994). *The 16PF Fifth Edition Technical Manual*. Institute for Personality and Ability Testing.

Davito, A. (1985). A review of the Myers-Briggs Type Indicator. In J. Mitchell (Ed.), *Ninth Mental Measurement Yearbook*. University of Nebraska Press.

Furnham, A. (1996). The big five Vs the big four: The relationship between Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality.

Personality and Individual Differences, 21(2), 303-307. https://doi.org/10.1016/0191-8869(96)00033-5

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.

Golubovich, J., Lake, C. J., Anguiano-Carrasco, C., & Seybert, J. (2020). Measuring achievement striving via a situational judgment test: The value of additional context. *Journal of Work and Organizational Psychology, 36*(2), 157-167. https://doi.org/10.5093/jwop2020a15

Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., Litman, L. (2022). Evaluating CloudResearch's Approved Group as a solution for problematic data quality on Mturk. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01999-x

Hauser, D. J., Paolacci, G., & Chandler, J. J. (2019). Common concerns with MTurk as a participant pool: Evidence and solutions. In F. R. Kardes, P. M. Herr, & N. Schwarz (Eds.), *Handbook of research methods in consumer psychology*. Routledge.

Kirk, R. E. (1996). Practical Significance: A Concept Whose Time Has Come. *Educational and Psychological Measurement, 56*(5), 746–759. https://doi:10.1177/0013164496056005002

Litman, L., Moss, A., Rosenzweig, C., & Robinson, J. (2021). *Reply to "MTurk, Prolific or panels? Choosing the right audience for online research."* https://doi.org/10.2139/ssrn.3775075

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151–159. https://doi.org/10.1037/h0026141

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

Noland, J. (1975). *Personalysis Reliability and Validity*. [Unpublished manuscript]. Personalysis Corporation.

Noland, J. (2005). *The development of Personalysis*. Unpublished manuscript. Personalysis Corporation.

Smith, N. A., Sabat, I., E., Martinez, L. R., Weaver, K., Xu, S. (2015). A Convenient Solution: Using MTurk To Sample From Hard-To-Reach Populations. *Industrial and Organizational Psychology, 8*(2), 220–228. https://doi:10.1017/iop.2015.29

Steckler, A., & Mcleroy, K. R. (2008). The importance of external validity. *American Journal of Public Health, 98*(1), 9–10. https://doi.org/10.2105/AJPH.2007.126847

Swift, V., & Peterson, J. B. (2019). Contextualization as a means to improve the predictive validity of personality models. *Personality and Individual Differences, 144,* 153–163. https://doi.org/10.1016/j.paid.2019.03.007

Thompson, M., & Borello, G. M. (1986). Construct Validity of the Myers-Briggs Type Indicator. *Educational and Psychological Measurement, 46*, 745-752.

Zhang, B., & Gearhart, S. (2020). Collecting Online Survey Data: A Comparison of Data Quality among a Commercial Panel & MTurk. *Survey Practice 13*(1). https://doi.org/10.29115/SP-2020-0015

Appendices

MBTI Convergent and Discriminant Validity Survey Demographics

Table 5

MBTI 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	7%	8%
Black or African American	30	10%	12%
Two or more races	13	4%	n/a
White or Caucasian	194	66%	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.

16PF Convergent and Discriminant Validity Survey Demographics

Table 6

16PF 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.