The Science Behind the Predictive Index Behavioral Assessment™

Table of Contents

About The Predictive Index

Serving more than 8,000 clients across 142 countries and delivering solutions in 70 languages, The Predictive Index offers a proven methodology that allows businesses to understand the factors driving their workforce. Through a unique blend of scientific assessments, ground-breaking software, highly-rated management training, and professional consulting from some of the world's best workplace behavior experts, The Predictive Index helps clients overcome their most complex business challenges.

Scientific validation and a 60-year proven track record have shown that business challenges big and small are no match for our unique approach to client education and knowledge transfer, which ensures swift adoption, direct return on investment, and high impact on performance metrics. Each year, 2.5 million PI Behavioral Assessments are administered and over 7,000 professionals are trained in our workshops.

Our assessments and recommended practices have been developed in a manner consistent with all critical standards and guidelines. These standards are designed to provide a framework for determining the proper use of assessments and other selection procedures, as well as to prevent discriminatory employment practices.

The Predictive Index assessments and recommended practices comply with standards and guidelines set by:

- Equal Employment Opportunity Commission (EEOC)
- Uniform Guidelines on Employee Selection Procedures (1978)
- American Psychological Association (APA)
- Society for Industrial and Organizational Psychology (SIOP)
- International Test Commission (ITC)

What is The Predictive Index Behavioral Assessment?

The original version of the PI Behavioral Assessment was developed and introduced by Arnold S. Daniels in 1955. The PI Behavioral Assessment is a theory-based self-report measure of normal adult personality that was built and validated extensively and exclusively for use within occupational and organizational populations.

- The PI Behavioral Assessment is untimed, takes approximately six minutes to complete, and employs a free-choice (as opposed to forced-choice) response format.
- Individuals are presented with two questions, each being followed by a listing of descriptive adjectives. Each response list contains the same 86 words with each word being associated with one of five Factors: Dominance, Extraversion, Patience, Formality and Objectivity (referred to as Factors A, B, C, D and E, respectively). Table 1 provides definitions of the Factors measured by the PI Behavioral Assessment.
- The first page of the assessment asks respondents to endorse adjectives that they feel describe the way they are expected to act by others (the *Self-Concept* domain).
- The second page asks respondents to endorse adjectives that they feel really describe them (the *Self* domain).
- The individual's responses are tallied to obtain scores for the appropriate Factor to produce the Self, Self-Concept and Synthesis (the average of Self and Self-Concept) patterns which are presented graphically.
- The patterns are then interpreted by users who have been trained through a Predictive Index workshop to utilize the PI Behavioral Assessment data for the selection, coaching, motivation and development of employees.

Table 1. PI Factor Definitions.

PI Factor	Description
DOMINANCE (A)	The degree to which an individual seeks to control his or her environment. Individuals who score high on this dimension tend to be independent, assertive and self-confident. Individuals who score low on this dimension tend to be agreeable, cooperative and accommodating.
EXTRAVERSION (B)	The degree to which an individual seeks social interaction with other people. Individuals who score high on this dimension tend to be outgoing, persuasive and socially-poised. Individuals who score low on this dimension tend to be serious, introspective and reserved.
PATIENCE (C)	The degree to which an individual seeks consistency and stability in his or her environment. Individuals who score high on this dimension tend to be patient, consistent and deliberate. Individuals who score low on this dimension tend to be fast-paced, urgent and intense.
FORMALITY (D)	The degree to which an individual seeks to conform to formal rules and structure. Individuals who score high on this dimension tend to be organized, precise and self-disciplined. Individuals who score low on this dimension tend to be informal, casual and uninhibited.
OBJECTIVITY (E)	The degree to which an individual prefers objectivity when processing information and making decisions. Individuals who score high on this dimension tend to be rational and logical and are typically influenced by facts and data. Individuals who score low on this dimension tend to be subjective and intuitive and are typically influenced by feelings and emotions.

The Construction of the PI Behavioral Assessment

The PI Behavioral Assessment has been in widespread commercial use since 1955, with revisions to the Assessment occurring in 1958, 1963, 1988, 1992 and 2016. While each successive iteration of the PI Behavioral Assessment has followed best practices in test construction, this document outlines the most recent revisions for the sake of brevity. The Predictive Index follows four primary steps when constructing and revising the PI Behavioral Assessment:

1. Content alignment
2. Establishing psychometric properties
3. Differential Item Functioning (DIF) analysis
4. Item finalization

Content Alignment Study

The purpose of content alignment studies is to quantify subject matter expert opinion about the alignment of existing and potential new words or items with the PI Behavioral Assessment Factors. In the latest revision, this involved the 86 existing words of the PI Behavioral Assessment and 140 pre-test words. Fifteen content experts, each with a doctorate in psychology, education, or a related field, participated in content alignment studies. All experts had extensive experience in assessment, education, and psychology, particularly as they apply to educational, work, and organizational environments. The content experts were asked to rate the degree to which each word aligns with each PI Behavioral Assessment Factor on a scale of 0 to 100. This information is used during the Item Finalization step (Foster et al, 2015).

Establishing Psychometric Properties

To obtain classical test theory item statistics, the new and existing words were administered to samples of actual test-takers. The Predictive Index collected 136,544 usable cases for analysis of the existing version of the assessment and approximately 10,000 cases for each of the 140 new pre-test words (Foster et al., 2015). Next, two classical test theory item-level statistics were computed for each studied word:

1. The proportion of people endorsing each word (i.e., the item mean response as items that are scored as one are endorsed and those scored as zero are not endorsed).
2. The correlation of the item with the Factor scores, or *item-total correlation* (ITC).

To obtain Factor scores, responses to the Self and Self-Concept items for each Factor (A through E) from the existing Form were submitted to exploratory factor analysis (EFA) with maximum likelihood extraction. A single-Factor (i.e., common Factor) solution was requested for each of these ten analyses, and Factor scores were saved so that ITCs could be subsequently calculated. The 140 pre-test words were included in each analysis (Foster et al., 2015). Item endorsement and ITC values for each Factor were evaluated during the Item Finalization step.

Differential Item Functioning

The final set of statistics used in the revision process was concerned with psychometric bias, which is also known as *differential item functioning* (DIF). DIF was examined for sex (male/female), ethnicity (majority/minority according to EEOC guidelines), and age (<40/40+ years). The resulting standardized difference statistic were interpreted using Cohen's traditional effect size guidelines (Cohen, 1988). Items with DIF effect sizes of 0.30 or greater were flagged for possible deletion from the final PI Behavioral Assessment Form V (Foster et al., 2015).

Item Finalization

Item Finalization involved evaluating the results of the content alignment study, the endorsement and ITC statistics, and the DIF effect sizes to arrive at a final version of the PI Behavioral Assessment. Below are the steps that led to the final assessment version (Foster et al., 2015).

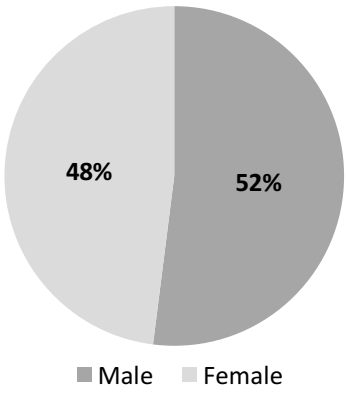
1. All words (both the current and the pre-test words) were sorted by content alignment ratings. Words were flagged for removal when they had low ratings for all five Factors or high ratings for multiple Factor designations (i.e., content experts on average agreed that the word could represent multiple Factors).
2. Using the psychometric data, items were removed when they had an extreme endorsement rate (i.e., endorsed by too few respondents [<10 percent] or too many respondents [>90 percent]).
3. Items were sorted by their ITC within each Factor and then compared against the Factors to which content experts believed they should belong from the alignment study. When an item's ITC was low (ITC < 0.30) within the content experts' recommended Factor, it was flagged for possible elimination.
4. Items exhibiting high DIF effect sizes (DIF > 0.30) were flagged for possible elimination.
5. A finalized 86-item form was then created by removing items that were flagged for elimination. For many of the items, there were clear data supporting their removal. For a few items, different data sources conflicted (e.g., content experts vs. psychometric data), and it was necessary to use judgment in the final decision.
6. A confirmatory factor analysis (CFA) was conducted to check the factor structure of the new form against the previous revision of the assessment. The results confirmed that the new items correlated with their intended factors.

Once the final form was completed, the assessment was normed and additional construct validation studies were conducted. Information about the norms and the construct validation studies is provided in the following sections.

PI Behavioral Assessment Norms

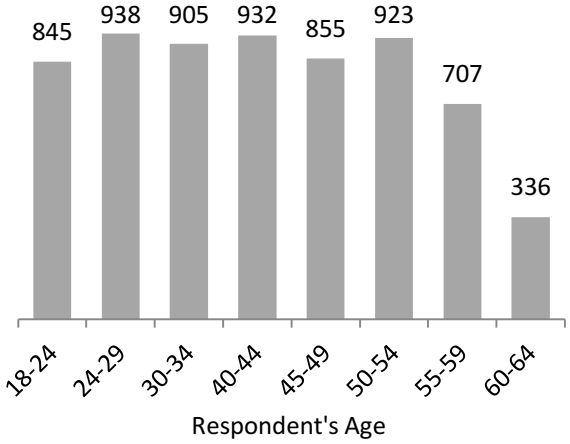
Norms are critical to the calculation of pattern scores, which are the basis of reporting and interpretation for end users. The PI Behavioral Assessment Norm is based on a sample of 9,645 of people. Of the sample's 7,658 respondents who reported their gender, 52% were male and 48% were female (see Figure 1). The average age of the 7,336 respondents who reported age in the normative sample was 40.1 years and ranged between 18 years and 64 years (see Figure 2). The norm is built from data from people in over 129 countries, although 64% percent of the sample was from the U.S. The primary ethnic composition of these U.S. respondents was Caucasian (74%), followed by African American or Black (12%), Latino, Latina or Hispanic (8%), and Asian (2%).

Figure 1. Norm Group Gender Distribution.



(*n* = 7,658)

Figure 2. Norm Group Age Distribution.



(*n* = 7,336)

Reliability, Validity, and Fairness of the PI Behavioral Assessment

The PI Behavioral Assessment is a well-constructed psychometric assessment that meets or exceeds the guidelines of Society for Industrial/Organizational Psychology (SIOP), the International Test Commission (ITC), and the American Psychological Association (APA). There are three common standards on which an assessment should be evaluated:

- *Reliability* – The precision of the scores and their consistency across testing instances (e.g., Is an assessment of adequate precision for the intended use? Are scores consistent if someone takes the assessment more than once?)
- *Validity* – Evidence supporting the interpretations and uses cases for the assessment (e.g., Does an assessment measure what it is intended to measure? Does is it predict what it is supposed to predict?)
- *Fairness* – Assessment considerations for applicable participant characteristics that could interfere with the validity of score interpretations (e.g., Does the assessment measure members of the population the same way? Are there any risks of adverse impact when using this assessment?)

Reliability Evidence

Research shows conclusive proof that the PI Behavioral Assessment has adequate to strong reliability as a psychometric instrument. The evidence for these claims is determined through three different methods:

- Test-Retest Reliability
- Internal Consistency Reliability
- Standard Error of Measurement Research

Test-Retest Reliability

The most straightforward approach to estimating reliability is repeated measurements of the same person. The correlation between temporally separated scores is known as the *test-retest reliability* or *coefficient of stability*. The Predictive Index has conducted four separate test-retest studies (see Table 2) ranging from two weeks to six months, and the findings show acceptable to strong test-retest reliability across all Self Factors and scales (with one exception noted below). Studies 1 and 2 were conducted by Everton (1999), and studies 3 and 4 were conducted by Harris in 2009 and 2011, respectively (Harris, Tracy, & Fisher, 2014). Note that these studies were conducted prior to the inclusion of Factor E (Objectivity) as a uniquely measured Factor.

Table 2. Test-Retest Reliability Coefficients from Four Samples.

PI Self Factor	Study 1 Two Weeks	Study 2 Six Months	Study 3 Two Weeks	Study 4 Two Weeks
Dominance (A)	0.80	0.75	0.84	0.79
Extraversion (B)	0.71	0.80	0.77	0.81
Patience (C)	0.76	0.71	0.75	0.70
Formality (D)	0.80	0.57*	0.83	0.87

Note: Study 1: n = 77; Study 2: n = 58, Study 3: n = 61; Study 4: n = 44.
*Everton's (1999) finding for Factor D in Study 2 appears to be an anomaly or a data transcription error. For all other test-retest studies, Factor D was among the most reliable Factors.

Internal Consistency Reliability

Another common way to estimate reliability is by computing *internal consistency*. The most common measure of internal consistency reliability is the coefficient alpha. Coefficient alpha is a conservative estimate of reliability, meaning that the true reliability of the assessment is either equal to or greater than coefficient alpha. Coefficent alpha reflects the extent to which item scores on a measure covary. Three samples of varying sizes were used to estimate internal consistency reliabilities using raw factor scores (Foster et al., 2015). As can be seen in Table 3, all internal consistency reliabilities were 0.85 or higher, indicating that the Form V scores have sufficient precision for use in selection and assessment.

Table 3. Internal Consistency Reliability Estimates for Three Samples.

PI Factor	Self			Self-Concept			Synthesis		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
Dominance (A)	0.87	0.88	0.87	0.88	0.88	0.85	0.91	0.92	0.91
Extraversion (B)	0.88	0.88	0.87	0.87	0.87	0.85	0.92	0.92	0.92
Patience (C)	0.87	0.86	0.85	0.88	0.86	0.85	0.92	0.91	0.91
Formality (D)	0.85	0.86	0.85	0.86	0.86	0.86	0.91	0.92	0.92
Objectivity (E)	0.87	0.86	0.86	0.88	0.86	0.87	0.92	0.92	0.92
Response Level (M)	0.96	0.96	0.96	0.96	0.96	0.96	0.98	0.98	0.98

Note: Sample 1: *n* = 1,023; Sample 2: *n* = 5,573, Sample 3: *n* = 4,072.

Standard Error of Measurement

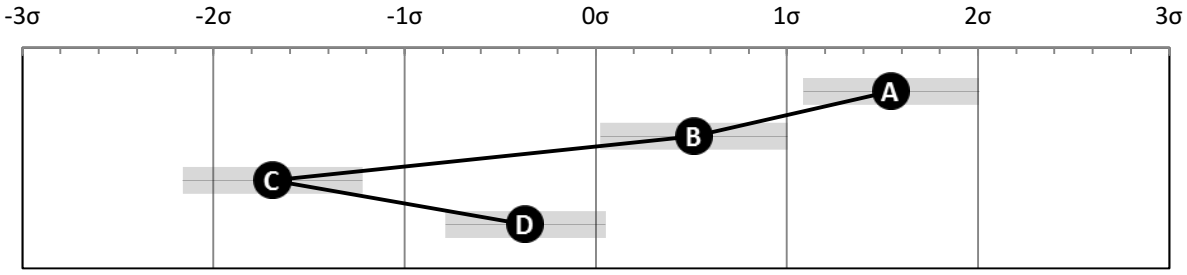
Finally, The Predictive Index has conducted internal consistency and *standard error of measurement* (SEM) estimates based on the Pattern Factor scores from the Norm Group sample (Fossey, 2016). Pattern Factor scores are used for graphic representations of PI Behavioral Assessment results in hiring and development reports. As such, they are normative transformations of the raw Factor scores (i.e., z-scores). Coefficient alpha was used to estimate the SEM values, meaning that the SEM values reported are conservative estimates: the true SEM for the instrument is equal to or less than these values. Table 4 shows the reliability estimates and the SEM for each Pattern Factor score. For all Factor scores, the SEM was less than 0.5σ.

Table 4. Reliability and Standard Error of Measurement (SEM) Using Norm Group Factor Scores.

PI Factor	Self		Self-Concept		Synthesis	
	Alpha	SEM(σ)	Alpha	SEM(σ)	Alpha	SEM(σ)
Dominance (A)	0.79	0.46	0.74	0.46	0.83	0.36
Extraversion (B)	0.73	0.49	0.73	0.49	0.81	0.35
Patience (C)	0.78	0.47	0.75	0.47	0.82	0.35
Formality (D)	0.83	0.42	0.82	0.42	0.86	0.31
Objectivity (E)	0.85	0.48	0.81	0.48	0.88	0.37

Interpretation of the PI Behavioral Assessment scores is typically done at a pattern level, thus the specific Factor scores do not matter as much as their relative placement to each other. The average absolute score differences between Factor scores for the Norm Group ranged from 1.1σ (between Factors A and B) to 1.7σ (between factors D and E). Thus, the impact of an SEM less than 0.5σ is unlikely to have an impact on the pattern-level interpretations of the scores, as the relative positions of the Factor scores are not likely to switch based only on the small SEM. Figure 3 provides an example of the ±1 SEM ranges around a PI Behavioral Assessment pattern.

Figure 3. Graphic Representation of PI Behavioral Assessment SEM.



Validity Evidence

The *validity evidence* of an assessment refers to its ability to:

1. Measure the psychological constructs that it is supposed to measure. This is called *construct validity*.
2. Predict important outcomes, such as job performance. This is called *criterion validity*.

The following sections summarize the research that has provided evidence that meets or exceeds professional guidelines for both kinds of validity. It is important to note that the PI takes only six minutes on average to complete. Given the brevity of the assessment, the evidence below is quite impressive.

Construct Validation

The Predictive Index has conducted multiple construct validation studies over its history, all of which provide solid evidence that the PI Behavioral Assessment Factors measure the psychological constructs that they are designed to measure. Table 5 provides selected findings from correlational studies against the NEO PI-R (Harris, Tracy, & Fisher, 2014), the 16PF (Everton, 1999), and scales from the International Personality Item Pool (IPIP) (Foster et al, 2015). Full results of convergent and divergent evidence are available in the referenced studies. Note that these studies were conducted prior to the inclusion of Factor E (Objectivity) as a uniquely measured Factor.

Table 5. PI Behavioral Assessment Construct Validation Evidence.

PI Factor	Instrument	Scale	*n*	*r*
Dominance (A)	IPIP	Assertiveness	1,023	0.50
	IPIP	Domineering	1,023	0.31
	16PF	Independence	103	0.47
	16PF	Dominance	103	0.46
	NEO PI-R	Agreeableness	186	-0.61

Table 5 (Continued). PI Behavioral Assessment Construct Validation Evidence.

PI Factor	Instrument	Scale	*n*	*r*
Extraversion (B)	IPIP	Extraversion	1,023	0.59
	IPIP	Sociability	1,023	0.54
	16PF	Social Boldness	103	0.46
	NEO PI-R	Extroversion	186	0.63
Patience (C)	IPIP	Stability	1,023	0.24
	16PF	Openness to Change	103	-0.32
	16PF	Dominance	103	-0.47
	16PF	Tension	103	-0.35
Formality (D)	IPIP	Methodicalness	1,023	0.43
	IPIP	Conscientiousness	1,023	0.37
	16PF	Perfectionism	103	0.37
	16PF	Independence	103	-0.63
	16PF	Self-Control	103	0.42
	NEO PI-R	Conscientiousness	186	0.61

Criterion-Related Validity

The Predictive Index has conducted hundreds of validity studies demonstrating that the PI Behavioral Assessment predicts job performance, thereby supporting the validity of its use for making talent decisions. Evidence of the criterion-validity of the PI Behavioral Assessment is presented in two ways:

1. Practical Validity – Evidence from The Predictive Index Validity Vault™.
2. Meta-Analytic Research – Evidence obtained through meta-analysis to identify generalizable relationships of the PI Behavioral Assessment to job performance across different performance criteria and job types.

Practical Validity Evidence

Over the last 60 years, The Predictive Index has conducted nearly 500 criterion validity studies. In 2015, a research program was established to collect and organize all client validation studies going back to 1992, when the fourth version of the PI Behavioral Assessment was released. The research program led to the construction of an extensive archive of criterion-validation studies that is called the PI Validity Vault. The PI Validity Vault represents practical validity, in that there are a staggering number of relationships between the PI Behavioral Assessment and performance across jobs, industries, and countries. Below are abridged details:

As of April, 2016, the Validity Vault included:

- 328 unique criterion validation studies.
- 7,566 correlational analyses conducted between PI Behavioral Assessment Factors and performance criteria.
- 111 unique job roles (based on O*NET codes) with the five most frequent being:
    - 115 Sales Roles
    - 39 Customer Service Roles
    - 33 Manager Roles
    - 20 Call Center Roles
    - 14 Teller Roles
- 11 different industries, with Retail, Finance/Insurance, Professional Firms, and Healthcare being the most common.
- PI Behavioral Assessment scores from over 25,000 working adults.
- 4,596 significant correlations between PI Behavioral Assessment results and various performance criteria.
    - Significant correlations were found in 94% of studies.
    - Half of these studies found moderate to high correlations (0.40 correlations or higher).

Meta-Analytic Validation Evidence

Three meta-analyses have been conducted, yielding support for the PI Behavioral Assessment's ability to predict sales performance, tenure, counterproductivity, and performance in management jobs, social jobs, and call center operation jobs. *Meta-analysis* is a statistical methodology designed to combine the results of many individual, independently conducted empirical studies into a single result or outcome (Hunter & Schmidt, 1990; Raju et al., 1991). The logic underlying meta-analysis is that researchers can arrive at a more accurate or generalizable conclusion about a particular topic by combining or aggregating the findings of a wide crossection of many studies that address the topic, instead of relying on the results obtained in a single study, which is susceptible to sampling fluctuations.

Study 1

The first meta-analysis was conducted in 2008 (Harris et al., 2014) and examined 57 predictive and concurrent criterion validity studies conducted from 2003 to 2007. The cumulative sample size for this meta-analysis was 5,765. The average sample size per study was 101 employees, with the lowest sample size being 32 and the highest being 431. The studies comprising the meta-analysis were conducted on behalf of 51 different organizations (including 21 Fortune 500 clients) drawn from 20 different industry classifications (e.g., "financial services") and 15 different occupational classifications (e.g., "customer service"). Fourteen of the 57 studies were conducted with managerial-level jobs, with the remainder focused on individual contributor roles such as bank tellers, customer service associates, truck drivers, machine operators, etc. The outcomes studied were job performance and job tenure.

Finally, the purpose of this study was to establish the generalizability and strength of the relationship to job performance (i.e., whether the PI Behavioral Assessment Factors related to job performance) and not directionality. For example, one might expect low Factor D score to be predictive of performance in sales, but high Factor D scores to be predictive of performance in administrative roles. If these findings are averaged, the resulting coefficient would be near zero, even though strong bi-directional relations exist. As such, the absolute values of all validity coefficients were computed before conducting the meta-analysis in order to avoid strong negative and positive results cancelling each other out.

The observed correlations between Self Factor scales and outcomes were sample weighted and then averaged to produce the meta-analytic estimates shown in Table 6. As can be seen in the table, the average observed correlations ranged from 0.11 to 0.19. These estimates are largely consistent with past meta-analytic investigations involving personality measures (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991), with the average validities being in the 0.10 to 0.20 range.

Table 6. Study 1 Meta-Analytic Validity Coefficients for Self Factor Scores.

PI Self Factor	Overall Job Performance		Tenure	
	Mean	SD	Mean	SD
Dominance (A)	0.17	0.09	0.13	0.08
Extraversion (B)	0.18	0.10	0.15	0.11
Patience (C)	0.18	0.10	0.14	0.08
Formality (D)	0.19	0.10	0.14	0.10
Objectivity (E)*	0.18	0.10	0.11	0.08
Response Level (M)*	0.18	0.12	0.14	0.09

*Measured at the Synthesis level.

Study 2

The second meta-analysis was conducted in 2014 (Foster et al., 2015) and involved 20 Predictive Index client organizations and 2,324 employees. The average sample size per study was 110 employees, with the lowest sample size being 45 and the highest being 421. The data were collected from multiple Predictive Index clients during 2007-2014 and covered a range of positions and industries.

In that meta-analysis, performance behaviors were grouped into different outcome categories. The two items that had the most meaningful findings were sales, performance, and counter-productivity. Examples of sales task performance (SP) measures are net sales, total sales, and percent sales numbers. The majority of jobs for this criterion were sales associates, sales specialists, and retail consultants. Aspects of managerial performance that involved customer acquisition or revenue growth also were classified as sales performance (e.g., net gain of patients in the medical practice). Counterproductivity, or counterproductive work behavior (CWB), is considered to be undesirable in that it represents employee behavior that is contrary to the organization's legitimate interests and is intended to harm the organization (Rotundo & Sackett, 2002). Examples of counterproductivity measures include attendance problems and the number of job-related errors.

Table 7 presents meta-analytic results for the four primary Factors (A through D) and the two outcomes, sales performance and counterproductivity. Mean validities are shown for raw scores as well as for pattern scores. As expected, Dominance (A) and Extraversion (B) scores were positively associated with sales performance, with average uncorrected correlations ranging between 0.07 and 0.12. The Extraversion Factor had a positive correlation with counterproductivity (0.17 - 0.19). For within-Factor scores, Patience (C) and Formality (D) had negative average observed correlations with counterproductivity and sales performance. If unreliability and range restriction corrections were performed, the observed validities might increase, which would match the published personality meta-analyses results for corrected correlations.

Table 7. Study 2 Meta-Analytic Validity Coefficients for Self Factor Scores (Raw and Pattern Scores).

PI Self Factor	Outcome	Raw Scores			Pattern Scores		
		Mean	Min	Max	Mean	Min	Max
Dominance (A)	Sales performance	0.10	-0.23	0.36	0.12	-0.02	0.35
	Counterproductivity	0.08	-0.14	0.35	-0.05	-0.21	0.12
Extraversion (B)	Sales performance	0.07	-0.29	0.27	0.10	-0.14	0.31
	Counterproductivity	0.17	-0.07	0.36	0.19	0.14	0.34
Patience (C)	Sales performance	-0.05	-0.37	0.24	-0.14	-0.40	0.09
	Counterproductivity	0.06	-0.26	0.12	-0.08	-0.35	0.06
Formality (D)	Sales performance	-0.04	-0.32	0.24	-0.11	-0.37	0.21
	Counterproductivity	0.07	-0.17	0.21	-0.11	-0.20	0.03

Table 8 presents meta-analytic results for the six pairwise combinations of the four primary Factors (A though D) and two outcomes, sales performance and counterproductivity. In Table 8, all of the validities coefficients are in the expected direction.

Table 8. Study 2 Meta-Analytic Validity Coefficients for Factor Combination Pattern Scores.

PI Self Factor Combination	Outcome	Expected Direction	Combination Scores		
			Mean	Min	Max
Task-Oriented (A>B)	Sales performance	n/a	0.01	-0.12	0.30
	Counterproductivity	Negative	-0.16	-0.03	0.37
Proactive (A>C)	Sales performance	Positive	0.14	-0.07	0.42
	Counterproductivity	n/a	-0.01	-0.10	0.40
Comfortable with Risk (A>D)	Sales performance	Positive	0.14	-0.17	0.34
	Counterproductivity	n/a	0.03	-0.28	0.20
Quick to Connect to People (B>C)	Sales performance	Positive	0.14	-0.12	0.30
	Counterproductivity	n/a	0.15	-0.03	0.37
Informal (B>D)	Sales performance	n/a	0.11	-0.07	0.42
	Counterproductivity	Positive	0.18	-0.10	0.40
Casual with Rules (C>D)	Sales performance	n/a	-0.03	-0.17	0.34
	Counterproductivity	n/a	0.03	-0.28	0.20

Study 3

In the third meta-analytic study, PI Behavioral Assessment data from 1,104 employees across 14 organizations and 26 different jobs were compared to supervisor performance ratings in four categories: overall job performance, compliance, cooperation, and initiative (Foster et al, 2015).

To facilitate the interpretation of criterion validity results, jobs with managerial responsibilities (e.g., project manager) were grouped into a *management* category. Jobs that involved direct physical interactions with customers (e.g., registered nurse) were grouped into a *social* category. Jobs with predominantly structured sales responsibilities (e.g., retail sales consultant) were grouped into a *sales service*, and jobs that involved solving customer problems over the telephone or online (e.g., senior web advisor) were grouped into a *call center operator* category.

After dividing jobs into four job families, meta-analytic validity estimates were obtained by averaging the sample-weighted corrected correlations between Factor scores (and their combinations) with job performance ratings. In addition, all observed correlations were corrected for supervisor rating unreliability.

Table 9 provides a high-level review of the most meaningful findings, with with coefficients in bold representing validity coefficients higher than 0.15 or lower than -0.15. A thorough summary of this work is provided in Foster et al. (2015), with a briefer presentation here.

The results showed that different Factors and combinations predicted performance for the different job families. This confirms the importance of using the PI Behavioral Assessment correctly. Different Factors have different relationships with job performance based on the behavioral demands of a job.

Table 9. Study 3 Meta-Analytic Results.

Job Family	Performance Criteria	Factors				Factor Combinations					
		A	B	C	D	A>B	A>C	A>D	B>C	B>D	C>D
Manager	Compliance	0.07	**-0.16**	-0.09	**0.19**	**0.17**	0.09	-0.07	-0.03	**-0.20**	**-0.20**
	Initiative	**0.28**	-0.09	**-0.15**	-0.05	**0.28**	**0.24**	**0.20**	0.05	-0.02	-0.07
Social Jobs	Performance	-0.12	-0.02	**0.15**	0.03	**0.20**	-0.12	-0.09	-0.11	-0.01	0.07
	Compliance	**-0.18**	-0.08	**0.20**	0.06	**0.16**	**-0.18**	-0.14	**-0.18**	-0.06	0.08
	Cooperation	**-0.16**	0.05	0.08	0.02	**-0.15**	-0.13	-0.09	-0.03	0.03	0.03
Call Center	Cooperation	0.11	0.13	-0.06	**-0.15**	-0.06	0.09	**0.16**	0.12	**0.15**	0.07
Structured Sales	Overall	-0.03	0.00	-0.14	**0.17**	-0.01	0.06	-0.13	0.08	-0.09	-0.22
	Compliance	**-0.16**	**-0.17**	0.04	**0.28**	-0.01	-0.11	**-0.27**	-0.13	**-0.25**	-0.14
	Cooperation	**-0.15**	0.13	-0.02	0.07	**-0.21**	-0.06	-0.13	0.08	0.03	-0.05
	Initiative	-0.06	-0.04	-0.06	**0.17**	-0.01	0.00	**-0.15**	0.01	-0.12	**-0.16**

Fairness

In employment settings, large differences in average scale scores across demographic groups can result in lower rates of selection of ethnic minorities, women, or older applicants. When these lower rates occur in the United States, it is called *adverse impact*. Adverse impact is obviously undesirable. To date, there is no evidence to indicate that the inclusion of the PI Behavioral Assessment in a company's personnel selection system, either in a compensatory or "multiple-hurdle" selection model, results in adverse impact against any protected class. When examining adverse impact via the four-fifths rule, Adverse Impact (AI) ratios for large-scale studies of the PI Behavioral Assessment ranged 0.86 to 1.25. In fact, in 60 years, there has never been a successful legal challenge involving the PI Behavioral Assessment. In addition, The Predictive Index has run the following studies to demonstrate that the use of its assessment tool does not result in adverse impact:

1. Wolman (1991) ran a study to determine whether men and women tended to score differently on the PI Behavioral Assessment, and whether African-Americans, Hispanics, and Caucasians tended to score differently. His analyses showed that neither gender nor race was significantly related to PI Behavioral Assessment scores.
2. In a more recent study by Harris (2004), the PI Behavioral Assessment was analyzed to determine whether it produces adverse impact based on age. The study showed that for all PI Behavioral Assessment Factors, there was no signficant difference between people over age 40 (the protected class) and people under age 40, confirming similar findings initially obtained by Everton (1998).
3. In a 2008 banking industry study of 347 employees working in a variety of jobs (e.g., teller, branch manager, loan officer), gender and race accounted for less than 2% and 3% of the variability, respectively, in PI Behavioral Assessment Factor scores (Harris, Tracy, & Fisher, 2014).
4. DIF analysis was conducted to look for item-level response bias based on gender, race/ethnicity, and age, and items with DIF effect sizes of 0.30 or higher were excluded from consideration for the construction of the PI Behavioral Assessment (Foster et al., 2015).

Both the DIF procedures used in the construction of the PI Behavioral Assessment and the results of multiple research studies indicates that the PI Behavioral Assessment is age-, gender- and race-neutral, and we believe that the inclusion of a well-validated personality assessment such as the PI Behavioral Assessment in a company's personnel selection system may lead to a more demographically diverse workforce.

Note: When requested by the client, The Predictive Index Science Team will partner with PI clients to examine adverse impact and other selection-related statistics. Custom research studies, including pass-rate and adverse impact analyses, can also be conducted for PI clients, designed in full accordance with Uniform Guidelines and other professional standards.

Conclusions

This document summarizes the science behind the PI Behavioral Assessment, which has been evolving scientifically since its inception in 1955. There is significant scientific evidence to show that the tool is a well-constructed, thoroughly validated assessment that can be used for workplace decision-making. Even more impressive, and often overlooked, is the fact that the instrument takes only six minutes to complete on average, yet yields such strong, accurate results. Below are key summary points in this document:

1. The PI Behavioral Assessment is constructed following best test development practices, including the use of content analysis with subject matter experts, psychometric evaluation, factor analysis of the constructs, and DIF to reduce the likelihood of unfairness (Foster et al., 2015).
2. The PI Behavioral Assessment has adequate to strong test-retest reliability and internal consistency (Everton, 1999; Harris, Tracy, & Fisher, 2014). The standard error of measurement is below 0.5σ for all Factors (Fossey, 2016).
3. Extensive construct validation work has compared the PI Behavioral Assessment Factors to relevant scales on the NEO PI-R, 16PF, and the IPIP scales. The results support the intended interpretations for the PI Behavioral Assessment Factors (Everton, 1999; Harris, Tracy, & Fisher, 2014; Foster et al., 2015).
4. The PI Behavioral Assessment has a 60-year history of criterion validation evidence that shows the PI Behavioral Assessment does predict job performance. The more recent Validity Vault work provides substantial practical evidence that the PI Behavioral Assessment predicts important workplace outcomes across jobs, industries, and performance criteria.
5. Three separate meta-analytic studies have been conducted. Each study adds a different angle of support for the generalizability of the tool for predicting job performance and for the proposition that the relation between Factors and performance is largely based on the behavioral demands of the role (Harris, Tracy, & Fisher, 2014; Foster et al., 2015). Furthermore, the results are consistent with past meta-analytic investigations involving personality measures (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991).
6. The PI Behavioral Assessment is a fair assessment. Multiple studies have demonstrated that PI Factor scores will not differ significantly by age, race/ethnicity, or gender, and DIF analysis is used to screen out adjectives that might result in response bias against a protected group (Wolman, 1991; Everton, 1998; Harris, 2004; Harris, Tracy, & Fisher, 2014; Foster et al., 2015).

References

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

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions in job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum Associates.

Everton, W. (1999). *A further investigation of the construct validity of The Predictive Index®*. (Validity Report No. 1999.11.01). Wellesley Hills, MA: Predictive Index, LLC (Formerly Praendex, Inc.)

Fossey, A. (2016). *Using sigma scores to estimate reliability and the standard error of measurement for Form V of the PI Behavioral Assessment* [technical document]. Westwood, MA: The Predictive Index.

Foster, D., Maynes, D., Miller, J., Chernyshenko, O. S., Mead, A., & Drasgow, F. (2015). *Meta-analysis report of The Predictive Index (Form IV)*. Wellesley Hills, MA: PI Worldwide.

Harris, T. C., Tracy, A. J., & Fisher, G. G. (2014). *Predictive Index® technical overview*. Wellesley Hills, MA: PI Worldwide.

Harris, T.C. (2004). *The impact of age on Predictive Index scores: An update*. Wellesley Hills, MA: PI Worldwide.

Hunter, J. E., & Schmidt, F. L. (1990). *Method of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76*, 432-446.

Tett, R.P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.

Uniform Guidelines (1978). *Uniform guidelines on employee selection procedures*. Washington, DC: Equal Employment Opportunity Commission (EEOC).

Wolman, R. N. (1991). *Predictive Index factors in relation to race, sex, and job levels*. Wellesley Hills, MA: PI Worldwide.