



Development of an Empirically-Based Short Form Personality Assessment

Heather Hayes, Camber Corporation

Brandon Ferrell, Hogan Assessment Systems

Jonathan Huck, University of Oklahoma

Blaine Gaddis, Hogan Assessment Systems

This paper was accepted as a SIOP poster for the 2017 conference.

Abstract

Short form personality assessments are nothing new, but many existing forms are constructed with an emphasis on internal reliability rather than predictive utility. We developed a short form using an empirical approach that simultaneously optimizes reliability and criterion validity. We also provide evidence supporting the utility of our short form.

Introduction

In recent years, short form versions of self-report and survey assessments have become increasingly popular in organizations (Stanton, Sinar, Balzer, & Smith, 2002). There are several advantages to short form assessments, such as the efficiency with which researchers can measure organizationally-relevant constructs and the speed of producing outputs useful for making selection and other administrative decisions. The demand for short forms has only accelerated with the increasing popularity of online and mobile assessment, including the use of smart phones and other handheld devices (Arthur, Doverspike, Muñoz, Taylor, & Carr, 2014) where programming is more complex and large numbers of items may cause problems in the response process (Peytchev & Hill, 2010).

Despite these positive features, there are a number of pitfalls and caveats associated with developing and using short forms. Thus, the process by which one is constructed is delicate and requires thorough quality testing. The goal of our research was to develop a short form version of a well-known personality assessment via a method that addresses concerns associated with short form development and use. Specifically, we used a strictly empirical process that optimizes both reliability and validity associated with the use and interpretation of scores obtained from the newly constructed short form. We now turn to issues and concerns, raised by researchers, surrounding short form construction.

Construction of Short Forms: Issues and Concerns

As short form versions of standard-length assessments have increased in popularity and availability, professionals have devoted more attention to the advantages and disadvantages of these forms. Some advantages of short forms are clear and agreed upon by many researchers and practitioners – namely, efficiency, convenience, and reduced fatigue among test-takers (e.g., Stanton et al., 2002). However, short forms are not without vocal critics who argue against their development and use (Smith & McCarthy, 1995). Other researchers recognize both the potential rewards and threats associated with short forms and recommend specific guidelines for developing and evaluating the adequacy of short form assessments (e.g., Smith, McCarthy, & Anderson, 2000; Widaman, Little, Preacher, & Sawalani, 2011).

The general concern with short form assessments is that they do not live up to the quality of the original version from which they were derived. Specifically, researchers voice concerns over the methods used to select items from the original instrument to construct a short form and the construct coverage of the resulting subsets of items. The two most prominent concerns with short forms are the reliability and validity inherent in short form score use and interpretations, and critics of short forms agree that evidence for both is required in short form scores (Smith et al., 2000; Widaman et al., 2011).

Reliability refers to precision of measurement (Lord & Novick, 1968) and consistency in scores across time, observers/raters, or among items intended to measure a unidimensional construct (i.e., coefficient alpha; Cronbach, 1951; McDonald, 1999). In fact, coefficient alpha represents the most common criterion for constructing short forms (Widaman et al., 2011). Specifically, researchers can use internal consistency to identify a subset of items from the original instrument that are most highly correlated with one another. Although internal consistency is crucial for representing the intended construct and providing a ceiling for validity due to measurement error and associated type II error rates (McDonald, 1999), the primary concern over its predominance in guiding short form construction is the attenuation paradox associated with sacrificing breadth of construct representation for unidimensionality (Loevinger, 1954; Smith et al., 2000; Widaman et al., 2011).

Validity can take many forms (Messick, 1995), all of which ultimately address the fundamental question of how well assessment scores reflect the intended theoretical latent construct (e.g., McDonald, 1999). Moreover, criterion-related validity is crucial because it concerns the intended purpose of the assessment – explaining and predicting non-test behavior. Thus, researchers have argued that validity should drive short form construction because it not only captures the essence of a construct, but also what it does – i.e., how well it accomplishes its goal of predicting non-test behavior or performance (e.g., John & DeSoto, 2007; Widaman et al., 2011). According to socio-analytic theory, personality is a social construct (e.g., Hogan, 1983, 1996; Wiggins, 1979). People present themselves in a certain way in order to manage the observer (i.e., scoring key; whoever sees the scores) impression of him or her, ultimately affecting the test-taker’s reputation (as evaluated by the “observer”). In other words, in personality assessment, scores should reflect the test-taker’s reputation as described by others. The implication of this theory is that one can understand and predict work behavior because people seek acceptance and status in the workplace, and their behavior and self-presentations reflect these goals. In other words, the construct underlying the measurement score is the behavior being predicted; there is no entity or construct outside of what we do and how we think with respect to our social context.

We describe these properties because at the most fundamental level, psychometrically adequate short forms cannot be derived from the original instruments that, themselves, lack proper reliability and validity evidence. Unless researchers begin with a foundation of such evidence, short forms based on poor quality instruments will bear out the cliché “garbage in, garbage out”. With the importance of the original instrument in mind, we used a well-known and extensively validated personality assessment as our marker instrument, and began our empirical process to develop a short form solution.

Method

Sample

Our research sample consisted of 4,918 adults employed across five job families: (a) managers and executives ($n = 1268$), (b) professionals ($n = 1041$), (c) sales ($n = 1081$), (d) customer service ($n = 662$), and (e) technicians and specialists ($n = 866$). To develop our short form, we used an exploratory sample including 35% of the total sample. This sample

included equal representation from each job family, 64% male and 36% female participants, and a mean participant age of 40 years old ($SD = 9.35$). In terms of racial/ethnic composition, the exploratory sample included 35% White, 11% Black/African-American, and 16% Hispanic/Latino participants; 37% of respondents answered “other” or did not indicate their racial/ethnic group. To evaluate the results of our short form, we used a cross-validation sample that also included equal representation from each job family, 62% male and 38% female participants, and a mean participant age of 40 years old ($SD = 8.75$). In terms of racial/ethnic composition, the cross-validation sample included 44% White, 16% Black/African-American, 25% Hispanic/Latino participants; 15% of respondents answered “other” or did not indicate their racial/ethnic group.

Assessment

The Hogan Personality Inventory (HPI; R. Hogan & Hogan, 2007) is a well-known measure examining “bright-side” personality, and was the first such inventory specifically developed for occupational contexts with working adults. It includes seven scales (see Table 1) that align with the Five-Factor Model (FFM; cf. Digman, 1990; Goldberg, 1992; McCrae & Costa, 1987) of personality. The seven HPI scales align with the Five-Factor Model, but with two important caveats. Specifically, Surgency or Extraversion from the FFM is split into separate Ambition and Sociability scales with the HPI, as a person may be goal-oriented and introverted or extraverted but not driven. The HPI also splits FFM Intellect or Openness to Experience into separate Inquisitive and Learning Approach scales because a person may be naturally curious but not value education or vice versa. With those two exceptions, the HPI scales align 1:1 with FFM dimensions – HPI Adjustment links to FFM Emotional Stability, HPI Interpersonal Sensitivity links to FFM Agreeableness, and HPI Prudence links to FFM Conscientiousness. Factor analysis results support this hierarchical structure, with HPI facets loading onto seven dimensions and those seven dimensions loading onto the appropriate higher-order factor.

The seven HPI scales are based on a socio-analytic theory of personality, which emphasizes the importance of interpersonal psychology (e.g., Hogan, 1983, 1996; Wiggins, 1979). This perspective emphasizes that our behavior is driven by competing motivations to get along with others to gain acceptance, and to get ahead of others to secure status and power (Wiggins & Trapnell, 1996). The HPI has been extensively validated since its inception, being included in over 500 criterion-related validity studies across all industry sectors and occupations covered by the US Department of Labor. Using this socio-analytic perspective, Hogan and Holland (2003) found that when predictors and criteria are conceptually aligned, meta-analytic validity estimates for HPI scales exceed those found in prior atheoretical meta-analyses. Specifically, the authors found that as performance criteria moved from general (i.e., overall job performance) to specific job criteria, all HPI scales more precisely predicted relevant performance outcomes. In fact, when they aligned criteria by personality construct, the estimated true validities of the HPI scales ranged between .25 (Learning Approach) and .43 (Adjustment) (Hogan & Holland, 2003).

The technical manual for the HPI (Hogan & Hogan, 2007) provides extensive evidence on the adequacy of the instrument in terms of structural psychometrics, reliability, validity, and factor structure. As such, the HPI shows ample evidence of being a high-quality marker

instrument on which to base short forms. The 206 self-report items included in the full assessment also make the HPI an excellent candidate for constructing short forms to maximize prediction in the age of mobile assessment.

Analysis

Using the assessment and sample described above, we began the multi-stage process of developing a short form of the HPI. During the development stage, we used our exploratory sample to select a subset of items from each scale to maximize prediction of job performance, then supplemented those predictive items with additional items needed to maximize other psychometric properties such as reliability and factor structure. Specifically, for each scale we selected items to maximize prediction of job performance using a forward stepwise regression in which we iteratively regressed job performance onto an accumulating number of items, one item at a time, until the change in R-square for the inclusion of an additional item was no longer statistically significant.

We used a composite of supervisor ratings on critical aspects of job performance as aligned performance outcome measures. Performance ratings ranged from 1 (*Fails to Meet Expectations*) to 5 (*Exceeds Expectations*), with an average performance rating of 3.16 ($SD = 1.77$). We chose the number of items for each scale based on changes in R-square associated with the addition of an item to the predictive model. Our goal in this step was to retain the minimum number of items per scale that optimized performance prediction. Table 2 presents the number of items retained for each scale and the R-square for each short scale.

With short scales drafted to maximize prediction, we then sought to maximize internal consistency of our short scales using a forward stepwise regression in which we regressed the total scale score onto all items, one scale at a time. As with the previous analysis, we retained items when there was minimal change in R-square from adding the item to the predictive model. Our goal in this step was to retain the minimum number of items per scale to achieve optimal reliability. Table 3 presents the number of items retained for each scale and the R-square for each short scale. We added these items to the predictive items retained from the previous analysis, resulting in the final short form of the HPI. The short form includes 7 Adjustment items, 7 Ambition items, 5 Sociability items, 6 Interpersonal Sensitivity items, 7 Prudence items, 7 Inquisitive items, and 5 Learning Approach items for a total of 44 items.

Results

To document results for both the original instrument and our short form, we used a separate cross-validation sample to calculate reliability indices and validity estimates. By comparing results of our short form to the original instrument, we can determine the empirical quality of the short form relative to the original long-form assessment. Specifically, we compared internal consistencies and observed (i.e., uncorrected) criterion correlations for short and full scales from the HPI as indices of reliability and validity, respectively. Tables 4 and 5 present results from these analyses.

As seen in Table 4, our short form does evidence some loss in internal consistency relative to the original HPI. However, this finding is not surprising given the fact that the 44-item short form includes only 21% of the total 206 items from the full HPI. Nevertheless, internal consistencies for short scales remain adequate, ranging from .60 (Interpersonal Sensitivity, Prudence) to .72 (Ambition), compared to estimates ranging from .69 (Interpersonal Sensitivity) to .85 (Adjustment, Ambition) for full HPI scales with the same sample.

Despite the fact that the short form includes 79% fewer items than the full HPI, criterion-related validity evidence is encouraging. As shown in Table 5, short versions of all 7 HPI scales significantly predicted job performance in our cross-validation sample, compared to 6 standard-length HPI scales. Moreover, short and long versions of two HPI scales (Sociability, Learning Approach) were essentially equivalent in predicting performance, and short versions of three HPI scales (Adjustment, Prudence, Inquisitive) were more predictive of job performance in our cross-validation sample than long versions of the same scales in the full HPI. In fact, only two HPI scales (Ambition, Interpersonal Sensitivity) were less predictive in short forms than in the full HPI.

Finally, we tested the latent factor structure of our short form assessment using CFA in Mplus. Table 6 presents results from these analyses, comparing the fit of three models with incrementally added parameters. In the first model, we loaded items onto one of seven dimensions representing the seven HPI scales and allowed all factors to correlate with one another. In the second model, we loaded the original seven factors onto five higher-order factors representing the dimensions from the Five-Factor Model, though we constrained these higher-order factors as uncorrelated. In the final model, we repeated the analysis from our second model but allowed the five higher-order factors to correlate with one another. Although the dimensions of the FFM are theoretically orthogonal, there is substantial support for overlap in variance among these factors (e.g., Digman, 1997). The third model provided the best fit for the data, supporting the theoretical latent structure of the HPI and providing evidence for consistency with the full marker instrument in this regard.

Discussion

The goal of our research was to develop a high quality short form personality assessment based not on arbitrary decision points on numbers of items or a singular emphasis on internal reliability, but on more extensive, empirical research evidence. We gave top priority to the validity of each short scale, first selecting items that most strongly predicted job performance ratings. Next, we augmented these initial short scales by including additional items to ensure adequate reliability and internal factor structure. With the short form constructed, we tested the new instrument on an independent cross-validation sample to obtain evidence supporting the reliability and validity of the short form. In general, the short form maintained both reliability and validity. Although some reliability estimates did drop into the .60s, such estimates are still acceptable for such short scales (De Vellis, 1991). Moreover, short form reliabilities were not substantially lower than those for full scales; the degree to which these estimates dropped was proportionate to the reliability estimates for the full scales. Likewise, uncorrected validity estimates for our short scales showed increased prediction for three short scales relative to their full scale counterparts, with equivalent prediction for another two short scales and observed losses in prediction for only

two short scales. Given that our short form includes 79% fewer items than the original marker instrument, these reliability and validity results support the viability of our short form for projects where it may not be possible to use the full assessment.

Our short form development methodology differs from most in that criterion-related validity was our primary goal, with reliability and factor structure kept as secondary and tertiary goals, respectively (Smith et al., 2000; Widaman et al., 2011). However, our approach is not without its limitations, most notably construct breadth and coverage. Specifically, each scale in the full HPI includes a number of subscales. Despite the fact that our short form retains the scale structure of the full HPI, the small number of items per scale did not allow us to retain the subscale structure under each scale. As such, our short form is ideal for projects where users need a condensed assessment to provide scale-based information, but may not be appropriate for projects where users need in-depth information on specific facets under each scale. For the latter, we recommend original, full-length assessment versions.

Despite this limitation, our research demonstrates that researchers can use empirical rather than rational methods to develop short form assessments that retain reliability and validity with far fewer items than original instruments.

References

- Arthur, Jr., W., Doverspike, D., Muñoz, G. J., Taylor, J. E., & Carr, A. E. (2014). The use of mobile devices in high-stakes remotely delivered assessments and testing. *International Journal of Selection and Assessment, 22*, 2, 113-123.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- De Vellis, R. R. (1991). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246-1256.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review in Psychology, 41*, 417-440.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.
- Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.) *1983 Nebraska Symposium on Motivation* (pp. 55-89). Lincoln: University of Nebraska Press.
- Hogan, R. (1996). A socioanalytic perspective on the five-factor model. In J. S. Wiggins (Ed.), *The Five-Factor Model of Personality* (pp. 163-179). New York: Guilford Press.
- Hogan, R. & Hogan, J. (2007). *Hogan Personality Inventory manual* (3rd ed.) Tulsa, OK: Hogan Assessment Systems.
- Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*, 100-112.
- John, O. P. & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461-494). New York, NY: Guilford Press.
- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin, 51*, 493-504.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.

- McDonald, R. P. (1999). *Test theory*. Mahwah, NJ: Erlbaum.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Peytchev, A. & Hill, C. (2010). Experiments in mobile web survey design: Similarities to other modes and unique considerations. *Social Science Computer Review, 28*, 319-335.
- Smith, G. T. & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7*, 300-308.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102-111.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology, 55*, 167-194.
- Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary Data Analysis* (pp. 39-61), Washington, DC: American Psychological Association.
- Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395-412.
- Wiggins, J. S. & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspective* (pp. 88-162). New York, NY: Guilford.

Table 1. HPI Scales and Descriptions

HPI Scale	Measures the degree to which a person seems...
Adjustment	Calm and self-accepting or, conversely, self-critical and tense
Ambition	Socially self-confident, leader-like, competitive, and energetic
Sociability	To need and/or enjoy interacting with others
Interpersonal Sensitivity	Perceptive, tactful, and socially sensitive
Prudence	Conscientious, conforming, and dependable
Inquisitive	Bright, creative, and interested in intellectual matters
Learning Approach	To enjoy academic activities and value education for its own sake

Table 2. Initial Short Forms Constructed to Predict Job Performance

Scale	Items in Short Scale	R-square
Adjustment	4	.19
Ambition	3	.17
Sociability	2	.08
Interpersonal Sensitivity	2	.13
Prudence	4	.22
Inquisitive	3	.12
Learning Approach	2	.08

Table 3. Items Added to Initial Short Forms to Predict Total Scale Score

Scale	Number of items	R-square
Adjustment	3	.55
Ambition	4	.72
Sociability	3	.60
Interpersonal Sensitivity	4	.59
Prudence	3	.46
Inquisitive	4	.63
Learning Approach	3	.64

Table 4. Internal Consistency Reliabilities for Short & Long Form HPI Scales

Scale	Short Form	Long Form
Adjustment	.65	.85
Ambition	.72	.85
Sociability	.62	.84
Interpersonal Sensitivity	.60	.69
Prudence	.60	.71
Inquisitive	.68	.79
Learning Approach	.62	.72

Table 5. Observed Criterion-Related Validities for Short & Long Form HPI Scales

Scale	Short Form	Long Form
Adjustment	.14**	.10**
Ambition	.10**	.17**
Sociability	.12**	.12**
Interpersonal Sensitivity	.10**	.17**
Prudence	.24**	.08**
Inquisitive	.08**	-.03
Learning Approach	.07**	.08**

Table 6. Factor Structure and Comparison of Model Fit

Model	Factor Structure	RMSEA	CFI
1	7 factors (HPI Scales)	.07	.59
2	5 factors^ (FFM Dimensions)	.07	.69
3	5 factors^^ (FFM Dimensions)	.05	.79

Note: ^ Factors uncorrelated; ^^ Factors correlated.