

Running Head: GLOBAL VS. PRIMARY SALES

Predicting sales performance with the 16PF Questionnaire: Global versus primary scales

Alan D. Mead
Illinois Institute of Technology

Gary Kustis
Watterson, Aldridge, and Associates, Inc.

Author Note

Alan D. Mead, Institute of Psychology, Illinois Institute of Technology; Gary Kustis, Watterson, Aldridge, and Associates, Inc. Correspondence concerning this article should be addressed to Alan D. Mead, Institute of Psychology, Illinois Institute of Technology, 3101 South Dearborn/2nd floor, Chicago IL 60616. E-mail: amead@alanmead.org; mead@iit.edu

Abstract

Industrial psychologists continue to debate the bandwidth-fidelity dilemma in personality assessment. Do longer, broad scales predict better than shorter, more specific scales? In this study, the job-relatedness (cross-validated prediction of sales volume) of the global scales of the 16PF Questionnaire were compared to the more specific, primary scales. It was hypothesized and found that the more specific scales would have higher levels of validity. Thus, this study adds another data point to the debate (in favor of specific scales), extends the bandwidth-fidelity debate to objective criterion, and adds to the literature on the validity of the 16PF Questionnaire.

Press Paragraph

Personality is the “drill in ANWR” issue of industrial psychologists. We disagree, sometimes heatedly, about the predictiveness of personality scales and, sometimes, even about the precise definition of personality. The currently, generally accepted definition is that personality consists of five broad dimensions. However, we also have more specific scales and there has been some debate about which type, broad or specific, is better. This paper adds another data point to the debate, finding that the more specific scales of the 16PF Questionnaire predict sales volume better than broader scales. This paper also provides confirming evidence that previous research about subjective criteria also holds for objective criteria.

Predicting sales performance with the 16PF: Global versus primary scales

Since 1991, when twin meta-analyses demonstrated the validity of personality measures for employee selection (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991), personality assessments have become popular for employee selection. While there may be several reasons for the popularity of personality assessment, clearly the meta-analytic demonstration of robust, non-zero validity coefficients for job performance was a primary factor for the popularity. And the work of those meta-analysts relied upon a revolution in the personality literature. Goldberg (1992, 1993) and others popularized the finding that most personality traits could be grouped under one of five broad personality dimensions, the so-called “Big-Five” model (Digman & Inouye, 1986).

The broadness of the Big Five model was absolutely necessary for the Barrick (Barrick & Mount, 1991) and Tett (Tett, et al, 1991) studies, because it would have been difficult or impossible to categorize the many scales used in prior research into narrow dimensions. For example, the *Comrey Personality Scales* (1970) includes, among its 40 facet scales, similar scales such as “Acceptance of social order” and “Intolerance for nonconformity”. It would be very difficult, and very arbitrary, to categorize the large number of assessments used in past research into such narrow facets—impossible, of course, for broad assessments, such as extraversion. Thus, we suggest that the Big-Five model was instrumental to the current use of personality assessment in selection work.

Also, the broad scales of the Big Five seem, logically, to be broad like general intelligence, which is quite predictive of many different job-related criteria (Schmidt & Hunter, 1998). The broadness of personality also seemed to match the broadness of overall job performance, and thus it seemed logical that the broad factors of the Big Five would predict overall job performance better than more specific, narrower scales (Ones & Viswesvaran, 1996).

Another reason for preferring the Big Five scales over the more specific scales, is that there is general (sometimes begrudging) agreement on the five broad scales but little agreement about the specific scales (Roberts, Chernyshenko, Stark & Goldberg, 2005). For example, two popular instruments for applied work are the NEO (NEO-PI-R and NEO-PI-3; Costa & McCrae, 2008) and the 16PF (Cattell, Cattell, & Cattell, 1993), which have almost identical five-factor models (Saucier & Goldberg, 2003) but extremely different specific scales. The NEO has exactly six specific scales for each of the five broad scales for a total of 30 which are based on a more theoretical “top-down” approach (Child, 1998). The 16PF has fifteen specific personality scales that were derived empirically and have some cross-loadings in their relationships with the five broader factors (Cattell, 1994).

Other researchers have suggested that the more specific scales would be more useful and this has sometimes been described as the “Bandwidth-Fidelity” problem in personality testing (Ones & Viswesvaran, 1996; see Hogan & Roberts, 1996 for another perspective). Mershon and Gorsuch (1988) re-analyzed two dozen predictive studies, using a statistical estimate of shrinkage. They found substantially better prediction for the primary scales. Roberts and his colleagues (2005) also reported better factor analytic results when using facets of conscientiousness.

In the present study, we present an analysis of an instrument, the 16PF (Cattell, et al., 1993) that has a two-level structure. One can use the broad, global 16PF scales, which mirror the Big Five (Saucier & Goldberg, 2003) or the much more specific primary scales. Further, in the current study we

are predicting an objective criteria—the actual sales performance of salespeople. This is a relatively unusual type of criteria not extensively studied, even in prior bandwidth-fidelity research.

Mainly relying on the slightly greater tally of Bandwidth-Fidelity research favoring specific scales (Mershon & Gorsuch, 1988; Ashton, 1998; Roberts, et al., 2005) we hypothesized that the specific, primary factors of the 16PF would predict sales volume at a higher degree than the broader, global scales.

Method

Sample. The sample consisted of 245 sales people in a midwestern financial services firm. The data collection period extended over a period of three years. Approximately 58% of the sample were men and the mean age was 35.1. After deletion due to missing data, 170 individuals were included.

Measure. The 16PF Questionnaire (Cattell, et al, 1993) is a widely used, comprehensive measure of adult personality. In some form, it has been in applied use for over five decades (Boyle & Barton, 2008). The instrument consists of 15 specific personality scales and one general reasoning scale (hence the name). The fifteen personality scales are further grouped into five broad second-order or “global” scales similar to the big five (Saucier & Goldberg, 2003). The primary scales of the extraversion global are: Warmth (A); Liveliness (F); Social Boldness (H); Privatness (N; loading negatively); and Self-Reliance (Q2; loading negatively). The conscientiousness factors are: Liveliness (F; loading negatively); Rule Orientation (G); Abstractedness (M; loading negatively); and Perfectionism (Q3). In addition to the personality scales, the instrument also has a social desirability scale called Impression Management (IM; Conn & Rieke, 1994). Each of these scales consists of 11-12 items and have two-week test-retest reliabilities in the range of 0.77-0.87 (Conn, 1994).

Sales volume, the primary criterion in this study, was reported by the client organization. It consisted of the total amount of financial services sold by an individual over the two-year period of the study, adjusted for tenure differences within the study period.

Procedure. Salespeople were tested as candidates prior to being hired. The 16PF results were used to make recommendations regarding the candidate, rather than actually being used as a selection test with specific cut-off values. Thus, the selected candidates might be somewhat restricted in the ranges of their scores, but not as severely as would be the case with direct range restriction.

Results

K-fold cross-validation. The sample size (N=170) is relatively small for regression analysis. Further, it is well-known that regression analyses in small samples are subject to excessive shrinkage and that this issue is exacerbated by having more predictors. Thus, any analysis that compares a few broad variables to a number of specific variables without cross-validation might find results favoring the many specific variables simply because of capitalization on chance. However, the small sample size made simple split-sample cross-validation risky.

Our solution was to use K-fold cross-validation (Stone, 1977; Efron & Tibshirani, 1993) which is similar to double cross-validation. In our case, we split the sample into four randomly equivalent quarters and performed four complete analyses. In the first analysis, we withheld the first quarter of

our sample when fitting the regressions (describe below). Then we used the fitted regression equations to predict sales volume for the withheld individuals in the first quarter of the sample. We then returned the first quarter, withheld the second quarter, and repeated the analysis. We did this two more times for the remaining quarters.

In this way, we performed our calibration on almost the entire sample but we were left with independent (i.e., cross-validated) predicted sales volume values for all individuals in the dataset. We then simply correlated these predicted values with the actual sales volumes to produce cross-validated validity coefficients based on the entire sample. [Note that the cost to “having our cake and eating it too” is that no one was actually scored using a regression computed on the entire sample; this cross-validation methodology therefore probably slightly underestimates the cross-validities that would be found if the entire sample could be used to calibrate the regression models.]

Hierarchical regression analyses. Based on previous validities research (e.g., Barrick & Mount, 1991) and job requirements, it was decided that Extraversion, Impression Management, and Conscientiousness would be most related to performance. Consequently, only these scores were used. The Impression Management (IM) scale on the 16PF is a single validity scale intended to index the degree to which respondents are creating a positive self-report of their personality.

Parallel hierarchical regression analyses were used to regress sales volume onto sets of predictors. In the global scale analysis, first Extraversion was entered, then Impression Management (IM), and then Conscientiousness. In the primary scales analysis, all Extraversion scales (A, F, H, N, and Q2) were entered as one step, then in the second step IM was entered, then all conscientiousness primaries (G, M, Q3).

It is important to note that the steps described above were based on theoretical predictions about the predictiveness. The order in which the predictors are entered substantially determines the (lack of) predictiveness for the latter steps. Therefore, this analysis does not prove that conscientiousness is unrelated to sales volume (see Table 1, discussed below) but merely that conscientiousness did not add much predictive power to the previously entered variables.

Table 1 presents the primary results. Three models for two types of scales (globals vs. primaries) and presented for two validation methodologies (cross-validated, and not cross-validated). When the models are not cross-validated, each additional step in the modeling (which introduces additional predictors) increases the validity coefficient. The effect is more pronounced for the primaries, where many more predictors are being entered at steps 1 and 3. Again, this difference is likely to be at least partially produced *artificially* by the uneven numbers of predictors. On the right side, the cross-validated validity coefficients are presented. They are considerably smaller, representing the shrinkage attributable to the small calibration sample size and some additional shrinkage due to the nature of the K-fold cross-validation.

Notably, the global scales, shown in the top half of Table 1, have obviously smaller validity coefficients, as compared to the primary scales, shown at the bottom of Table 1. Global Extraversion is essentially unrelated to sales volume while the primary scales have a modest but useful degree of validity. Impression Management substantially increases the predictiveness of the model already containing the global extraversion, while it makes a modest contribution to the model employing the

primary scales. In both cases, conscientiousness seems to slightly decrease the validity. However, the residual primary scale coefficient (0.11) is still of a useful size.

The detrimental effect of adding conscientiousness was unexpected. In reviewing the regression output, the regression weights for some of the conscientiousness scales were in a counter-intuitive direction (e.g., predicting that rule-ignoring individuals would sell more). Probably these reflect a combination of a small sample and a weak relationship between sales performance and integrity during the first months of employment.

Discussion

The primary research question of this study was regarding the relative merits of using the global or primary scales for predicting sales performance. The results in Table 1 show a clear advantage for the more specific primary scales. To take the Table 1 results at face value, the primary scales provide a coefficient 43% larger than the largest of the global scales.

However, even more convincing to us is that the cross-validated coefficients of all three specific, primary scale models are of roughly the same useful level (i.e., 0.11 to 0.14) whereas only one of the models involving the global scales rises to this level. We believe that this phenomenon is a clear indication of the advantage of using a more granular level of predictors which allow a suitable model (i.e., a proper regression model in this case, an improper, theoretically-guided model in another) to include those aspects of the broad five factors that are more relevant. After all, why should the dimensionality of real-world criterion just *happen* to match the dimensionality of broad, global scales? Why not use more granular scales that can be individually weighted to achieve better results? We feel that our results answer this question in the direction of clearly favoring more specific scales.

The results in Table 1 also emphasize the importance of cross-validation. The results evaluated in the same sample used to calibrate the regression model produced results that seem to significantly overstate the predictive validity of the scales. Although the pattern of the main research question regarding specificity of the scales

One unique aspect of this paper is the use of an objective measure of performance. Clearly, there are few more overarching or more vital criteria of the performance of sales people than the actual dollar amount of their sales and this research extends the discussion on the bandwidth-fidelity issue to this type of criterion.

The use of an objective criterion (sales volume) probably also affected the absolute level of prediction. The best predictive validity coefficients were only modest and we feel that this may well be due to the lower reliability common to objective criteria (see Campbell, McCloy, Oppler, & Sager, 1993). We did not have a good estimate of the reliability of the objective criterion. If the criterion reliability were 0.50, then the corrected cross-validated results would range from 0.04 to 0.14 and 0.16 to 0.20 for the global and specific scales, respectively.

Finally, we believe that these results may say as much about the 16PF as about the specificity of the scales. As described in the introduction, the specific scales of various instruments are vary a great deal in the “narrowness” and similarity to other scales. The primary scales of the 16PF are not based on some a priori theory or a rational parsing of the Big Five but rather on extensive empirical analysis.

Raymond Cattell found and largely ignored the second-order model because he felt that the primary scales were much more useful (Cattell, 1996). We believe that the number and composition of these scales may be partially responsible for the present results; it remains to be seen whether similar results might be obtained with NEO (Costa & McCrea, 2008) or Comrey (1970) scales.

Limitations and future directions. Obviously, the smallish sample size is a limitation. We believe that our use of K-fold cross-validation removed the deleterious effects of capitalization on chance but replication on a much larger dataset would be not only confirm these results but better fix the absolute level observed.

Replication in other organizations, with other types of jobs and other kinds of objective criteria would all be obviously beneficial in establishing the robustness and generality of these findings.

We believe that researchers should pay attention to the type of instrument used in research. It would be interesting to compare these results with other instruments, as described above. The primary scales of the 16PF Questionnaire were created differently from other scales facet scales and the traditional 16PF usage favors the specific scales. For instruments, such as the NEO, which emphasize the five-factor model and which have very different facet scales, it is not clear that identical results would occur.

References

- Ashton, M.C. (1998). Personality and job performance: the importance of narrow traits. *Journal of Organizational Behavior*, 19(3), 289-303
- Barrick, M.R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1-26.
- Boyle, G.J., & Barton, K. (2008). Contribution of Cattellian personality instruments. In G.J. Boyle, G. Matthews, & D.H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment*, Vol. 2, (pp160-178). Thousand Oaks, CA: Sage.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1992). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel Selection in Organizations* (pp. 35-70). San Francisco, CA: Jossey Bass.
- Cattell, H.E.P. (1996). The original big five: A historical perspective. *European Review of Applied Psychology*, 46(1), 5-14.
- Cattell, H.E.P. (1994). Development of the 16PF fifth edition. In S.R. Conn & M.L. Rieke (Eds), *The 16PF fifth edition technical manual* (pp. 75-100). Champaign, IL: IPAT.
- Cattell, R.B., Cattell, A.K., & Cattell, H.E.P. (1993). *16PF Fifth Edition Questionnaire*. Champaign, IL: IPAT.
- Child, D. (1998). Some technical problems in the use of personality measures in occupational settings illustrated using the "Big-Five". In S. Shorrocks-Taylor (Ed.), *Directions in Educational Psychology*, (pp. 346-364). London: Whurr Publishing.

- Conn, S. R. (1994). Reliability and equivalency: Comparison of the 16PF fifth edition and fourth edition (form A). In S.R. Conn & M.L. Rieke (Eds), *The 16PF fifth edition technical manual* (pp. 75-100). Champaign, IL: IPAT.
- Conn, S. R. & Rieke, M. L. (1994). Response style indices. In S.R. Conn & M.L. Rieke (Eds), *The 16PF fifth edition technical manual* (pp. 53–73). Champaign, IL: IPAT.
- Costa, P.T., & McCrae, R.R. (2008). The revised NEO personality inventory (NEO-PI-R). In G.J. Boyle, G. Matthews, & D.H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment*, Vol. 2, (pp160-178). Thousand Oaks, CA: Sage.
- Comrey, A.L. (1970). *The Comrey Personality Scales*. San Diego, CA: Educational and Industrial Testing Service.
- Digman, J.M. & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology*, 50, 116-123.
- Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. London: Chapman & Hall.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26-34.
- Goldberg, L. (1992). The development of markers for the big five structure. *Psychological Assessment*, 4, 26–42.
- Hogan, J. & Roberts, B.W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, 17(6), 627-637.
- Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology*, 55, 675-680.
- Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729.
- Ones, D.S., & Viswesvaran, C. (1996). Bandwidth-Fidelity Dilemma in Personality Measurement for Personnel Selection. *Journal of Organizational Behavior*, 17(6), 609-626.
- Paunonen & Ashton (2001).
- Saucier, G. & Goldberg, L.R. (2003). The structure of personality attributes. In M.R. Barrick & A. M. Ryan (Eds.), *Personality and work: Reconsidering the role of personality*, (pp. 1-29). San Francisco: Josey-Bass.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Stone, M. (1977). Asymptotics for and against cross-validation. *Biometrika*, 64, 29-35.
- Roberts, B.W., Chernyshenko, O.S., Stark, S., & Goldberg, L.R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, 58, 103-139.

Tett, R.P., Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44(4), 703-742.

Table 1.
Hierarchical regression validities and cross-validities for two types of scales (global vs. primary)

Scales	Model	Calibration Sample Validity	Cross- Validity
Global	1 (Ex)	.152	.030
	2 (Ex, IM)	.201	.100
	3 (Ex, IM, Co)	.202	.088
Primary	1 (A, F, H, N, Q2)	.273	.119
	2 (A, F, H, N, Q2, IM)	.291	.143
	3 (A, F, H, N, Q2, IM, G, M, Q3)	.312	.111

Note: N=170. IM=Impression Management; Globals: Ex = Extraversion; Co=Conscientiousness; Primaries: A=Warmth, F=Liveliness, H=Social Boldness, N=Privateness, Q2=Self-Reliance, G=Rule Consciousness, M=Abstractedness, Q3=Perfectionism.

Note: With 168 degrees of freedom and a one-tailed test with alpha equal to 0.05, a simple correlation of 0.15 would be significantly different from a null hypothesis of zero correlation.