

# Criterion-Related Validity of an Innovative CAT-Based Personality Measure

**Robert J. Schneider**  
PDRI

**Richard A. McLellan and Tracy M. Kantrowitz**  
PreVisor, Inc.

**Janis S. Houston and Walter C. Borman**  
PDRI

*Presented at the CAT Research and Applications Around the World Poster Session,  
June 2, 2009*



## **Abstract**

This paper describes development and initial criterion-related validation of the PreVisor Computer Adaptive Personality Scales (PCAPS), a computerized adaptive testing-based personality measure that uses an ideal point IRT model based on forced-choice, paired-comparison responses. Based on results from a large consortium study, a composite of six PCAPS scales identified as relevant to the population of interest (first-line supervisors) had an estimated operational validity against an overall job performance criterion of  $\rho = .25$ . Uncorrected and corrected criterion-related validity results for each of the six PCAPS scales making up the composite are also reported. Because the PCAPS algorithm computes intermediate scale scores until a stopping rule is triggered, we were able to graph number of statement-pairs presented against criterion-related validities. Results showed generally monotonically increasing functions. However, asymptotic validity levels, or at least a reduction in the rate of increase in slope, were often reached after 5-7 statement-pairs were presented. In the case of the composite measure, there was some evidence that validities decreased after about six statement-pairs. A possible explanation for this is provided.

## **Acknowledgment**

**Presentation of this paper at the 2009 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC<sup>®</sup>.**

## **Copyright © 2009 by the Authors**

**All rights reserved. Permission is granted for non-commercial use.**

## **Citation**

**Schneider, R. J., McLellan, R. A., Kantrowitz, T. M., Houston, J. S., & Borman, W. C. (2009). Criterion-related validity of an innovative CAT-based personality measure. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved [date] from [www.psych.umn.edu/psylabs/CATCentral/](http://www.psych.umn.edu/psylabs/CATCentral/)**

## **Author Contact**

**Robert J. Schneider, PDRI, 650 3rd Ave S., Suite 1350, Minneapolis, MN 55402, U.S.A.  
Email: Robert.Schneider@PDRI.com**

## **Criterion-Related Validity of an Innovative CAT-Based Personality Measure**

This paper blends rigorous and innovative psychometric theory with a practical selection application. We used computerized adaptive testing (CAT) principles to develop a general assessment of normal adult personality, titled The PreVisor Computer Adaptive Personality Scales (PCAPS).

### **Origins of PCAPS**

PCAPS is based on programmatic research, beginning with that of Borman and his colleagues (Borman, Buck, Hanson, & Motowidlo, Stark, & Drasgow, 2001), who developed computerized adaptive rating scales (CARS) for evaluating job performance in the contextual or citizenship performance domain (Borman & Motowidlo, 1993; Motowidlo, Borman, & Schmit, 1997). Subsequently, CARS was expanded to measure the entire managerial performance domain, including task (Schneider, Goff, Anderson, & Borman, 2003). The CARS concept was initially transported into the personality domain in work conducted for the Navy (Houston, Borman, Farmer, & Bearden, 2005), which yielded the Navy Computer Adaptive Personality Scales (NCAPS). Given the promising CARS and NCAPS results, we sought to develop a commercial application of a personality assessment guided by our previous work. The PCAPS were designed to be predictive across a wide range of occupations and industries. Our intent was to build in flexibility to create composites of scales relevant to a variety of different work populations to accommodate the differing needs of our clients. To our knowledge, PCAPS represents the first commercial application of CAT methodology to the personality domain.

### **Psychometrics of PCAPS**

Nearly all of the major personality instruments currently in use have been constructed and scored using classical test theory methods generally attributed to Likert (1932). The most familiar item format presents a number of statements describing behaviors, attitudes or beliefs, and the examinee responds to each statement by indicating how well it describes him- or herself, using a response scale that is common across items. These methods are founded on dominance assumptions of the item response process (Stark, Chernyshenko, Drasgow, & Williams, 2006).

Briefly, the dominance model assumes that a person will tend to agree with a positively worded item if his/her trait level is more positive than that of the item. Similarly, a person will tend to agree with a negatively worded item if their trait level is more negative than that of the item. Clearly, items that are located at the middle of the continuum would not be informative under the dominance response process. Endorsing such an item would essentially be like saying “I agree that I am undifferentiated with regard to the trait being measured by this item.” In other words, the item would have no “direction” associated with it, and would be unscorable. In a scale constructed under dominance assumptions, only statements that reflect some non-neutral standing on the trait of interest have any information value. Thus, scales constructed in this manner tend to contain items that are more extreme, and the burden of test construction falls more heavily on creating enough non-neutral items that sample the behavioral domain of interest such that there will be variance in total scores within a population of examinees when item scores are summed to a scale score. The classical test theory techniques typically used for analysis and refinement of scales (e.g., factor analysis, discarding items with low item-total

correlation, etc.) serve to reinforce these outcomes. A clear implication of scale construction under dominance process assumptions is that measurement will be more precise for people who are located near the middle of the trait continuum. Such a scale will not be very effective in discriminating the trait levels of two persons who are both located far from the middle of the continuum, which could be of real importance depending on how scores are used.

One key aspect of item response theory (IRT) scoring models used under the dominance assumption is that item response functions are assumed to increase monotonically with trait level, across the entire range of trait levels. That is, the higher a person's standing on a trait of interest, the greater the likelihood of a "positive response" (i.e., endorsement of a positively worded statement, or selecting the correct answer on an ability test). Monotonicity makes sense in the domain of cognitive ability, where there is usually a demonstrably correct answer. For personality measurement, however, there is good reason to question dominance assumptions (Chernyshenko, Stark, Drasgow & Williams, 2007). There have been some investigations aimed at fitting personality data with IRT analyses within a dominance framework. Some of these have shown reasonably good fit, yet this could have resulted merely because the personality instruments studied had been developed within the dominance framework, and these methods lead to the rejection of items with non-monotonic response functions. Chernyshenko et al. (2007) make a compelling case that an IRT model that accommodates items with non-monotonic response functions might fit personality response data better than IRT models that assume the dominance response process, leading to greater flexibility in test construction and improved measurement precision in the personality domain.

PCAPS is based on ideal point assumptions of the response process, in contrast to the dominance process assumptions that have been implicit in the development and scoring of most existing personality measures (Stark et al., 2006). The ideal point paired-comparison approach has its roots in Thurstone's (1927) Law of Comparative Judgment. Thurstone conceived of using a paired-comparison procedure to scale stimuli on an interval scale. The idea was that if interval scale measurement could be achieved with a paired-comparison procedure, then measurement might be made more precise than that yielded by typical Likert-type personality scales, which arguably provide only ordinal level data. Stark and Drasgow (1998) developed an algorithm to implement this process based on Zinnes and Griggs' (1974) probabilistic unfolding model which, in turn, is based on (and extends) the work of Coombs (1950) and Thurstone (1927). Examinees select which of two statements representing different levels of a personality trait are more descriptive of them, and then the next pair of statements is selected using an updated trait level estimate based on their previous responses. Sequences of statement pairs are selected in a manner that maximizes item information at each step. Presentation of statement pairs for a given personality trait terminates when either a sufficiently low conditional standard error of measurement is reached, or the maximum desired number of statement pairs have been presented.

## **The Present Research**

In this paper, we briefly describe development of PCAPS, development of a composite of PCAPS scales for supervisory personnel, and initial criterion-related validity results for PCAPS. Because intermediate personality trait estimates are determined by a Bayesian estimation process based on the trait levels of the statements selected as most descriptive of examinees, we had a unique opportunity to investigate criterion-related validity levels associated with presentation of differing numbers of statement-pairs. If, for example, validity levels reach asymptotic levels after

the presentation of seven statement-pairs, that might be evidence that the algorithm should be modified to limit the number of statement-pairs presented to examinees to seven, rather than some larger number. This would make the measurement process more efficient, which is one of the benefits associated with CAT-based measurement.

## Method

### Development of PCAPS

We began the PCAPS development process by formulating a conceptual taxonomy of personality constructs, based on review of current personality theory and research, with particular attention to relevance for workplace applications. Thirteen dimensions were identified and defined, including Achievement Striving, Social Orientation, Cooperation, Composure, Adjusting to Change, Positive Self-Concept, Responsibility, Thoroughness, Sense of Duty, Self-Sufficiency, Learning Orientation, Directing and Influencing, and Innovation. Names and definitions of these constructs are shown in Table 1.

In order to maintain the integrity of the 13 PCAPS constructs throughout the process of creating and scaling statements, we identified facets within each scale to facilitate creation of statement pools that would have a representative balance of all aspects of the scale.

**Generation of item content.** For each scale, the ideal point paired-comparison approach requires a large pool of statements, each with a trait level parameter representing a specific location on the defined range of the trait ( $\theta$ ), and the statements must cover this range relatively evenly. Based on our previous research (e.g., Houston et al., 2005), it was estimated that approximately 150 to 200 statements would be needed for each of the 13 PCAPS scales, with a focus on populating the item bank with statements that evenly cover the trait distribution.

Ten senior-level I/O psychologists (with a mean of 12.3 years of experience;  $SD = 6.7$ ) were recruited to author the additional statements needed to populate the 13 PCAPS statement pools. A training workshop was conducted by a subset of the core project team to orient the statement authors to the task, familiarize the authors with the PCAPS taxonomy and the target population, the psychometric measurement model, and to distribute individualized statement-writing plans according to dimension and facet. The initial assignments for each author involved writing approximately 6-12 statements for each of two or three scales, distributed across the trait level range, to be reviewed by the core project team members who had been involved with the development of NCAPS. These statement reviewers then provided detailed feedback to each author about each statement, to reinforce the training and help calibrate the authors in terms of targeting their statements to specific trait levels. Following this feedback, each author was given their full writing assignment. In all, 2,679 statements were generated to represent all trait levels on the thirteen personality dimensions.

**Table 1. PCAPS Scales and Definitions**

| Scale                 | Definition   |
|-----------------------|--|
| Achievement Striving  | Setting and accomplishing challenging goals; taking satisfaction and pride in producing high quality work and excelling in one's own efforts; working hard, exerting effort, and persisting despite significant obstacles; competing with self and others.   |
| Social Orientation    | Seeking and enjoying social interaction; being friendly, sociable, warm and likeable; being pleasant, and easy to work with; being comfortable working with new people; thriving on social contact, such as working with others.   |
| Cooperation           | Valuing agreement, consensus, and cohesiveness among people; being receptive to others' attempts to build relationships; demonstrating sensitivity to others' needs and feelings; being understanding, helpful, and cooperative; accepting people readily and valuing connections with others; wanting to like other people. |
| Composure             | Maintaining composure and retaining the ability to think clearly and take effective action when confronted with stressful or negative situations; maintaining a positive outlook when criticized; putting aside worries and guilt.   |
| Adjusting to Change   | Easily and effectively handling change; demonstrating flexibility in approaching work; liking variety in work; working effectively with many different types of people in various situations and with differing constraints (e.g., high versus low autonomy).  |
| Positive Self-Concept | Believing in own abilities and skills; feeling competent and successful in multiple areas; remaining self-assured and optimistic even in the face of rejection.  |
| Responsibility        | Being accountable for own actions and their consequences; consistently honoring commitments and responsibilities; demonstrating reliability and planfulness; not becoming distracted or bored by routine tasks; not procrastinating even when tasks are unpleasant or unexciting.  |
| Thoroughness          | Meticulously keeping track of details without becoming overwhelmed by them; being exacting, precise, and accurate; spotting minor imperfections or errors and taking action to correct them; keeps belongings in order.  |
| Sense of Duty         | Accepting and complying with laws, regulations, rules, policies, and procedures; accepting authority; demonstrating honesty and trustworthiness; protecting sensitive or confidential information as appropriate.  |

**Table 1. PCAPS Scales and Definitions (continued)**

| Scale                     | Definition  |
|---------------------------|---|
| Self-Sufficiency          | Demonstrating willingness to make decisions independently; being self-sufficient and resourceful; exercising responsible autonomy; not being inappropriately dependent on others to get things done.  |
| Learning Orientation      | Being inquisitive; demonstrating a desire to increase own knowledge; enjoying the learning process and abstract thought; needing to understand how things work especially during the learning process; taking advantage of learning opportunities; applying learned material to new situations; learning from mistakes; taking advice; asking questions when unsure about something.  |
| Directing and Influencing | Persuading and negotiating effectively with others; influentially asserting ideas and thoughts; adeptly moving others to a decision or favorable outcome; effectively networking with others; coordinating individuals' efforts to accomplish work.   |
| Innovation                | Being innovative, creative, imaginative, and inventive; examining applied problems systematically and logically; taking an objective and purposeful approach to problem-solving; producing novel solutions to problems encountered on the job that are likely to be actionable and practical; using successful thinking strategies to solve problems, make decisions, creating unique ideas or work products; being open to multiple ideas and alternative modes of thinking. |

***Collection and analysis of trait level ratings.*** The trait level parameter of individual statements is central to the measurement model and adaptive functionality of PCAPS. The precision of the trait level estimates is crucial because the ideal point paired-comparison model derives item information based in part on the difference in the trait levels of each statement in a pair, and it is well known that the measurement error of a difference score is a compound of the measurement error of the individual elements upon which the difference is based.

Thirteen Internet-based rating surveys were created, one for each of the 13 PCAPS constructs. A total of 45 Master's and Ph.D. level I/O psychologists within PreVisor and PDRI with expertise in the domain of personality research were recruited to serve as trait raters. Sixteen of the raters were from PDRI, and most of those had previously participated in numerous trait rating processes identical to this one. Because of the PDRI raters' previous experience with this rating task, their responses were, to some degree, utilized as a benchmark for examining the quality of responses from other raters. Each rater provided responses for an average of 694.4 statements across 7.07 dimensions, and each statement was rated by an average of 22 raters. It should be noted that the 10 people who had served as statement authors also participated in the trait rating process, and the rating assignments were organized such that no rater was asked to rate a statement that he or she had written.

Two hundred fifty-one statements were ultimately eliminated, based on analyses of the rating data. Subsequently, additional qualitative and content review eliminated another nine statements. The mean value of  $r_{wg}$ , a measure of interrater agreement (James, Demaree, & Wolf, 1984, 1993), across the remaining statements was .92, and the 5th and 95th percentile values of  $r_{wg}$  were .86 and .96, respectively. Only two statements had an  $r_{wg}$  value below .80, and those were both in the upper .70s. As such, the interrater agreement of the trait level ratings for the items in the PCAPS pool was very good to excellent. The final PCAPS pool consisted of 2,519 statements.

## **Validity Sample**

PCAPS was administered to supervisory and managerial employees in a consortium of eight organizations participating in the validation of managerial assessments. This series of studies provided initial validity evidence for PCAPS. In the present research, we focus on the supervisor data. PCAPS was administered to 1,607 first-line supervisors in these eight organizations.

## **Criterion Data**

In the present research, the job performance criterion was an overall job performance composite, based on ratings from the examinees' immediate supervisors. This composite was formed by averaging across: (1) an average of 27 items linked to 10 job performance dimensions shown by job analysis to be important to supervisory and managerial positions, and (2) an average of seven general, global rating items. To compute these composites, items were standardized across the consortium data set. Composites were computed as the average  $z$  score of non-missing ratings.

Rating data were eliminated for an examinee if any of the following applied:

1. The rater had supervised the rated employee for less than four months.
2. The rater indicated that he or she was "not at all familiar" with the employee's job performance.
3. The length of supervision was four to five months, and frequency of observation was not at least weekly, *OR* the rater did not know the employee's performance at least "fairly well."
4. The length of supervision was six months or longer, and frequency of observation was not at least weekly *AND* the rater did not know the employee's performance at least "fairly well."

After applying these case-selection criteria, 1,109 of the 1,607 were retained for analysis.

## **Results and Discussion**

### **Formulation of a PCAPS Composite for Supervisors**

The present research dealt only with supervisors. As such, it was first necessary to identify a composite of scales relevant to first-line supervisors. To do this, we conducted a relative weight analysis (Johnson, 2000) to identify the relative importance of each predictor based on its proportionate contribution to  $R^2$ . This procedure controls for multicollinearity among predictors by considering the unique effect of each predictor as well as its effect when combined with the other predictors. In addition to results of the relative weight analysis, we also took into account

zero-order correlations between PCAPS scales and job performance, including the ability to predict facets of job performance that other scales could not. Six scales were identified: Achievement Striving, Thoroughness, Self-sufficiency, Directing and Influencing, Innovation, and Responsibility. Unit-weighted composites were formed by inversely weighting each scale according to its standard deviation and summing the weighted scale scores.

### Descriptive Statistics for PCAPS Scales

Descriptive statistics for the six PCAPS scales making up the supervisor composite are shown in Table 2.

**Table 2. Means, Standard Deviations, and Intercorrelations of Six PCAPS Scales in the Supervisor Composite**

| PCAPS scale                  | <i>M</i> | <i>SD</i> | 1   | 2   | 3   | 4   | 5   | 6  |
|------------------------------|----------|-----------|-----|-----|-----|-----|-----|----|
| 1. Achievement               | 6.07     | .66       | --  |     |     |     |     |    |
| 2. Responsibility            | 6.21     | .76       | .39 | --  |     |     |     |    |
| 3. Thoroughness              | 5.92     | .81       | .30 | .42 | --  |     |     |    |
| 4. Self-Sufficiency          | 6.43     | .58       | .19 | .12 | .05 | --  |     |    |
| 5. Directing and Influencing | 5.53     | .56       | .46 | .31 | .25 | .18 | --  |    |
| 6. Innovation                | 5.92     | .83       | .40 | .26 | .21 | .19 | .48 | -- |

*Note.*  $n = 1,607$ . Mean PCAPS scale scores can range from 2-8.

Mean scale scores ranged from 5.53 (Directing and Influencing) to 6.43 (Self-Sufficiency). Mean scores, therefore, ranged from approximately .75 to 2.0 standard deviations above the midpoint of the PCAPS scale (5.0). Intercorrelations between PCAPS scales ranged from .05 (between Thoroughness and Self-Sufficiency) and .48 (between Innovation and Directing and Influencing). The median intercorrelation was .26. In general, the pattern of intercorrelations was predictable. For example, Achievement, Responsibility, and Thoroughness were all moderately intercorrelated. This would be expected, given that they are all facets of Big-Five Conscientiousness. On the other hand, Self-Sufficiency was uncorrelated with Thoroughness, and there was no reason to expect a correlation between those two PCAPS scales.

### Validity Analyses

Validity coefficients were obtained by conducting a local meta-analysis (Oswald & McCloy, 2003). In addition to uncorrected validities, we also computed estimated operational validities (i.e., validities corrected for criterion unreliability only) of the PCAPS Supervisor composite, and its component scales, against our overall job performance criterion. The criterion reliability value used to make this correction was .60, slightly larger (i.e., more conservative) than the meta-analytic average of a single supervisor's rating reported by Viswesvaran, Ones, and Schmidt (1996). Validity results are shown in Table 3.

Estimated operational validity coefficients for the six PCAPS scales in the Supervisor composite ranged from .11 for Innovation to .23 for Achievement. The estimated operational validity coefficient for the Supervisor composite was .25.

**Table 3. Uncorrected and Corrected (Estimated Operational) Validity Coefficients Between Six PCAPS Scales in the Supervisor Composite and Overall Job Performance**

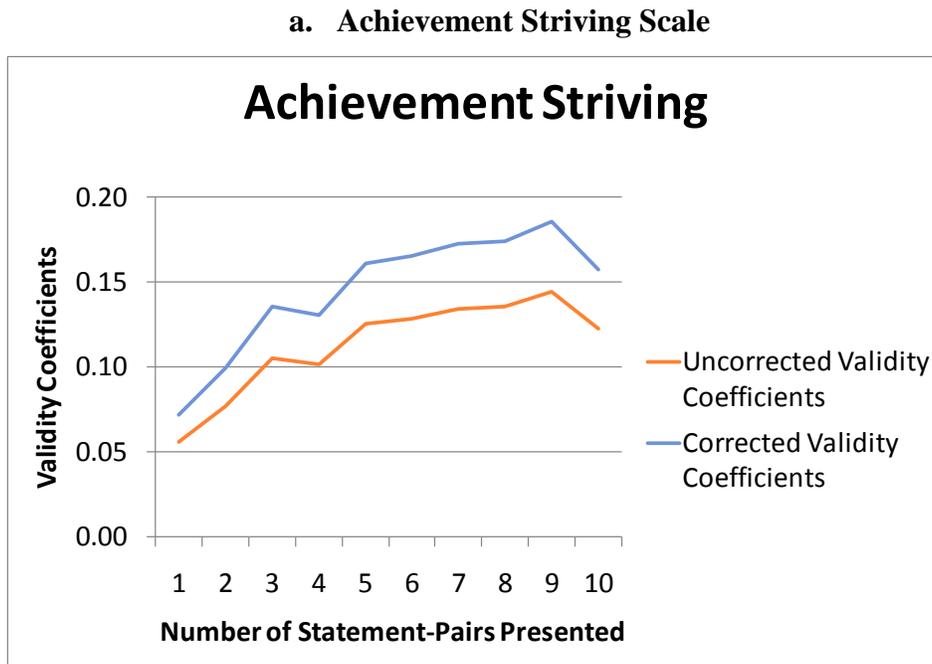
| PCAPS Scale/Composite     | Job Performance Composite |           |
|---------------------------|---------------------------|-----------|
|                           | Uncorrected               | Corrected |
| Achievement               | .17                       | .23       |
| Responsibility            | .14                       | .17       |
| Thoroughness              | .10                       | .13       |
| Self-Sufficiency          | .11                       | .14       |
| Directing and influencing | .13                       | .17       |
| Innovation                | .09                       | .11       |
| Supervisor composite      | .19                       | .25       |

*Note.*  $n = 1,099$  to  $1,109$ .

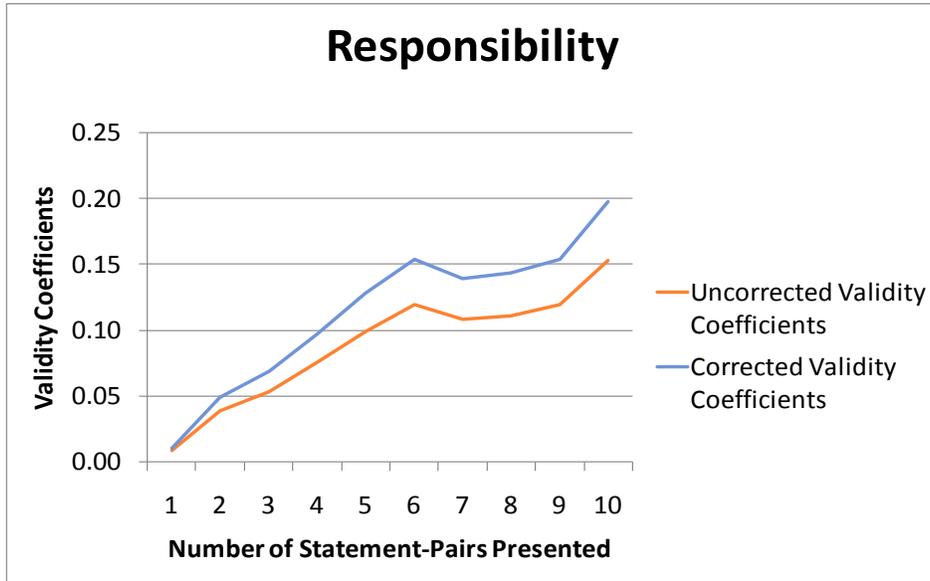
### Validities Associated With Differing Numbers of Statement-Pairs

Graphs showing validity coefficients associated with presentation of different numbers of statement-pairs are shown for each scale included in the PCAPS Supervisor composite, as well as for the Supervisor composite itself, are shown in Figure 1.

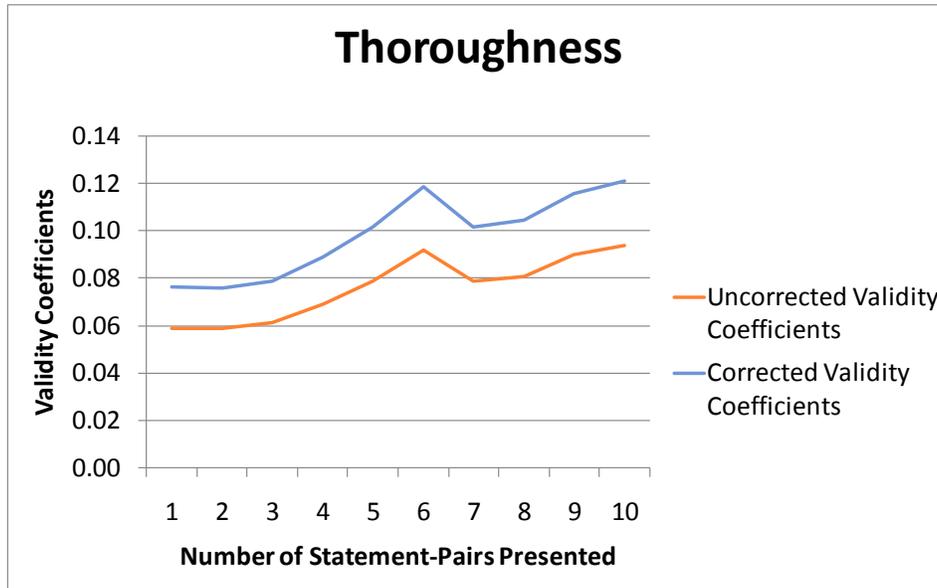
**Figure 1. Validity Coefficients as a Function of Number of Statement-Pairs Presented**



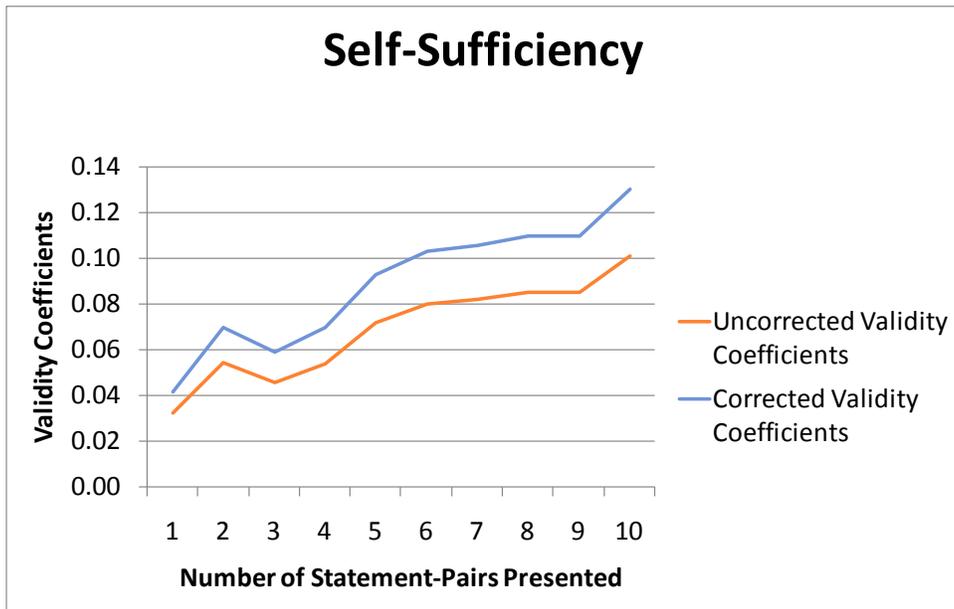
**b. Responsibility Scale**



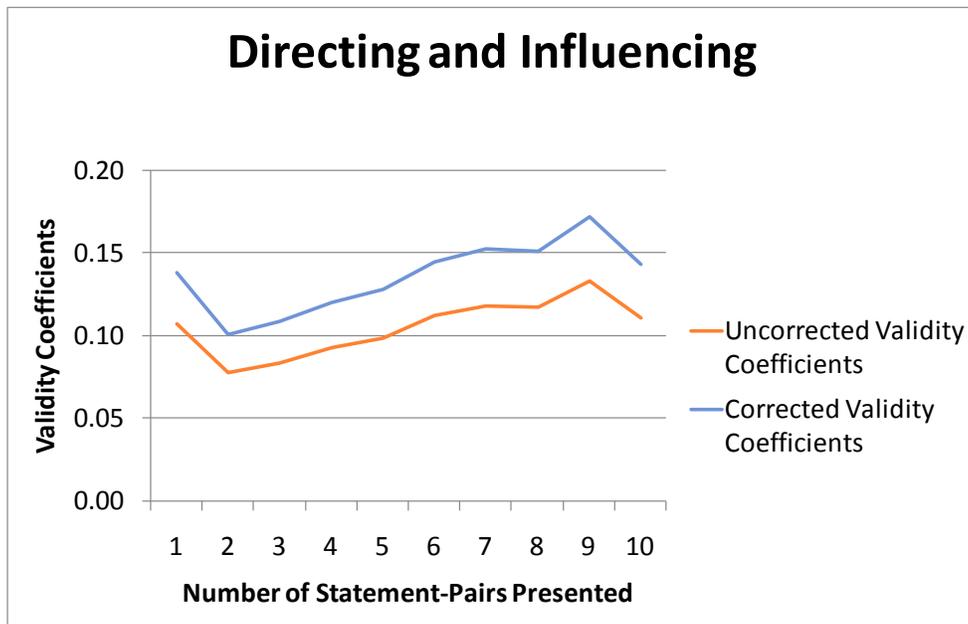
**c. Thoroughness Scale**



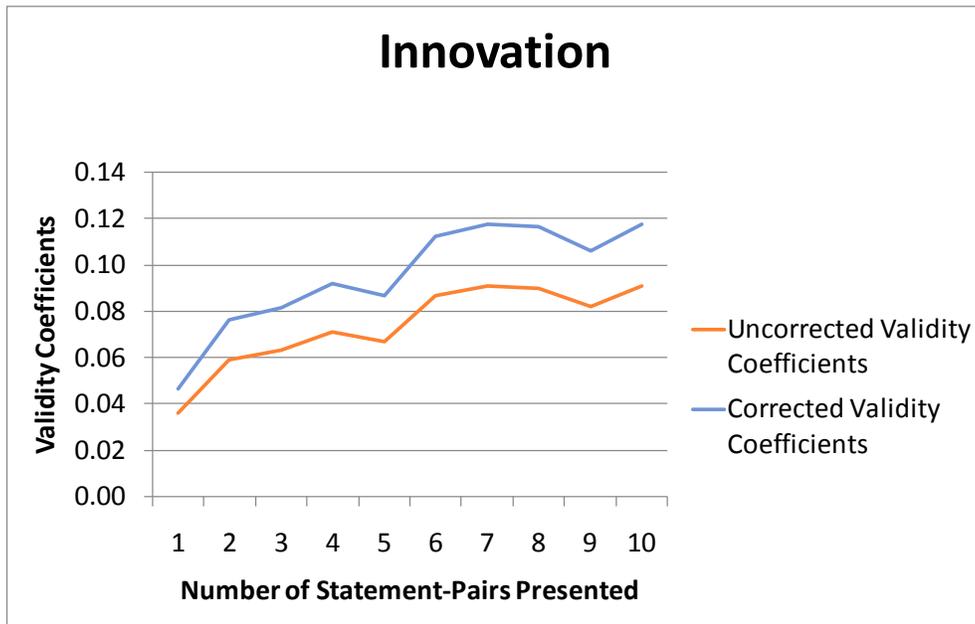
**d. Self-Sufficiency Scale**



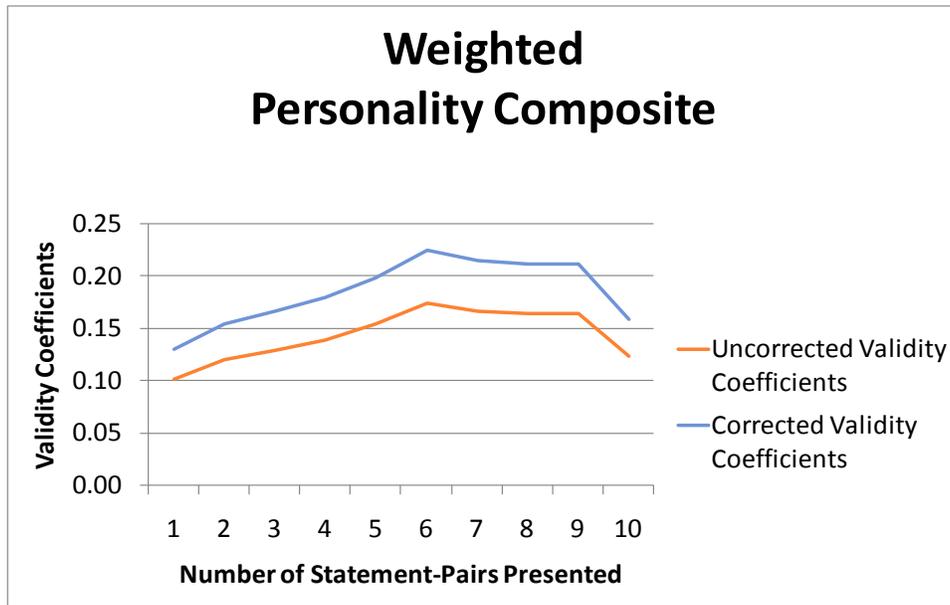
**e. Directing and Influencing Scale**



**f. Innovation Scale**



**g. Supervisor Composite**



The graphs in these figures are generally monotonically increasing, suggesting that greater validity might be achieved if more than 10 statement-pairs were presented. Interestingly, however, Achievement Striving, which had the highest validity of any of the six PCAPS scales in the Supervisor composite, shows signs of reaching asymptotic levels after as few as five statement-pairs. Of greatest interest were the graphs depicted in Figure 1g, which involved the

Supervisor composite. This is of interest because (1) it is the metric that would be used to make personnel decisions; (2) it is a broader trait, consisting of six PCAPS scales, and therefore comes closer to matching the overall job performance criterion in breadth; and (3) it is based on more information.

The Supervisor composite graphs look a great deal like the Achievement Striving graphs. It shows that the maximum validity coefficient is reached after six statement-pairs have been presented. It even shows some signs of reduction in validity after nine statement-pairs have been presented, though it is unclear whether this downward trend between nine and ten statement-pairs would continue. Even in the case of scales that do not appear to reach asymptotic levels prior to presentation of ten statement-pairs, the rate of increase in validity appears to slow after five to seven statement-pairs have been presented. This is true, for example, in the case of the Responsibility, Thoroughness, Self-Sufficiency, and Innovation scales.

It should be noted that the terminal validities depicted in Figures 1 are lower than those in Table 3. This is because the PCAPS stopping rules caused a terminal scale score to be reached for some examinees before they had been presented with 10 statement-pairs. It is also possible that examinees who are more certain of their trait levels require fewer statement-pairs to be presented before the stopping rules terminate presentation of statement-pairs for a given scale. That is, there may be subgroups differing in their level of traitedness (i.e., the extent to which they possess social-cognitive schemata representing the traits in question; see, e.g., Tellegen, 1988).

### **Summary and Conclusions**

This study evaluated a sophisticated and innovative CAT-based personality measure that uses an ideal point IRT model based on forced-choice, paired-comparison responses. Rigorous methods were used at every stage of PCAPS development, including delineation of the measurement space, creation and scaling of item content, engineering of the IRT-based adaptive functionality and scoring model, and the criterion-related validation of the PCAPS scales within a consortium of client research partner organizations.

PCAPS was tested on a large sample from a consortium of eight organizations. The initial validity results reported in this paper, for PCAPS scales relevant to first-line supervisors, provide criterion-related validity support, and reveal that as few as six statement-pairs may need to be presented to reach terminal validities for the scales in the Supervisor composite. While more research is clearly needed to replicate and extend these results, our initial results may support even greater efficiencies in test administration time, particularly at the composite level.

### **References**

- Borman, W. C., Buck, D. E., Hanson, M. A., Motowidlo, S. J., Stark, S., & Drasgow, F. (2001). An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology, 86*, 965-973.
- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

- Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19*, 88-106.
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review, 57*, 145-158.
- Houston, J. S., Borman, W. C., Farmer, W. L., & Bearden, R. M. (2005). *Development of the Enlisted Computer Adaptive Personality Scales (NCAPS), renamed Navy Computer Adaptive Personality Scales (NCAPS)* (Institute Report #503). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*, 85-98.
- James, L. R., Demaree, R. G., & Wolf, G. (1993).  $r_{wg}$ : An assessment of within-group interrater agreement. *Journal of Applied Psychology, 78*, 306-309.
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*, 1-19.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-53.
- Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10*, 71-83.
- Oswald, F. L., & McCloy, R. A. (2003). Meta-analysis and the art of the average. In K. Murphy (Ed.), *Validity generalization: A critical review* (pp. 311-338). Mahwah, NJ: Erlbaum.
- Schneider, R. J., Goff, M., Anderson, S., & Borman, W. C. (2003). Computerized adaptive rating scales for measuring managerial performance. *International Journal of Selection and Assessment, 11*, 237-246.
- Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91*, 25-39.
- Stark, S., & Drasgow, F. (1998, April). *Application of an IRT ideal point model to computer adaptive assessment of job performance*. Paper presented at the Society for Industrial and Organizational Psychology Annual Conference, Dallas, TX.
- Tellegen, A. (1988). The analysis of consistency in personality. *Journal of Personality, 56*, 621-663.
- Thurstone, L. L. (1927). Psychophysical analysis. *American Journal of Psychology, 38*, 368-389.
- Viswesvaran, C., Ones, D.S., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-560.
- Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika, 39*, 327-350.