

---

# Content Validation Is Fundamental for Optimizing the Criterion Validity of Personality Tests

---

THOMAS A. O'NEILL AND RICHARD D. GOFFIN  
*The University of Western Ontario*

ROBERT P. TETT  
*The University of Tulsa*

We do not dispute Murphy's (2009) arguments that content validity can add little to the criterion validity of selection tests that show strong positive manifolds. However, we disagree with his claim that use of some noncognitive tests, personality tests in particular, may not benefit from content validation in hiring situations. Referring to personality testing, Murphy suggests that "the evidence that content validity actually *does* matter in these contexts is relatively thin" (p. 462, emphasis in original). In contrast, we posit that *content validation has been strongly supported in so far as being associated with considerably larger criterion validity coefficients in personality testing*. In our view, content validation is the most direct way of identifying job-relevant personality traits and, accordingly, personality scales likely to demonstrate criterion validity.

Our goals in this commentary are twofold. First, we review evidence bearing on the extent to which personality trait

validities vary meaningfully across jobs in both magnitude and direction and whether experts can identify job-relevant traits a priori using content-oriented methods. Our review leads us to challenge Murphy's (2009) skepticism regarding the value of content validation applied to personality test use. Second, we link our review to existing scientific theory that describes *how* and *why* content validation can identify the job relatedness of personality traits and suggest that the value of content validation as applied to personality testing is extensive. We begin by discussing the nature of personality-performance relations.

## Selective Validity of Personality Trait Measures

Murphy's (2009) claim regarding the limits of content validation is based on the assumption of positive manifold, and where positive manifolds clearly exist, we support his reasoning. However, a critical characteristic of positive manifolds is that predictors are related to job performance *nonselectively*, that is, that different tests within a predictor domain are consistently correlated with a given criterion at similar magnitude and direction. To the contrary, research has shown that personality trait scale validities do vary across jobs (i.e., their

---

Correspondence concerning this article should be addressed to Thomas A. O'Neill.  
E-mail: toneill7@uwo.ca

Address: Department of Psychology, Social Science Center, The University of Western Ontario, 1151 Richmond St., London, Ontario, Canada N6A-5 C2

Thomas A. O'Neill and Richard D. Goffin, Department of Psychology, Social Science Center, The University of Western Ontario; Robert P. Tett, Department of Psychology, The University of Tulsa.

validity is situation specific) and that content validation can help identify especially job-relevant traits. We offer the following lines of supporting evidence.

Murphy and Shiarella (1997) demonstrated that the validity of selection tests varies considerably depending on how performance is defined—an argument we believe is especially relevant to personality testing. In addition, Murphy (1989) suggested that personality tests will be most criterion valid with respect to typical performance (e.g., maintenance stages) versus maximum performance (e.g., transition stages) because motivation is the key in the former case and personality traits are important motivational determinants. Thus, theoretical rationales support the argument that personality traits can be expected to show differential prediction across situations, suggesting that the predictive power of personality trait scores is selective.

In other research examining the selective validity of personality traits, Raymark, Schmit, and Guion (1997) presented a personality-based approach to job analysis in which experts familiar with several jobs were asked to rate the relevance of personality dimensions for performance in those jobs. The result was that expert ratings of content overlap between personality traits and job performance reliably differentiated 13 job families. Hogan and Holland's (2003) meta-analysis found that, when traits were thematically matched to criterion constructs, trait scale validities were higher than those when traits were not so matched. Dudley, Orvis, Lebiecki, and Cortina's (2006) meta-analysis revealed that facets of conscientiousness varied in the direction of their relations with job criteria. For example, the trait order was related to performance in managerial and customer service jobs in negative and positive directions, respectively. Hough's (1992) meta-analysis showed that personality trait validity differed greatly by job dimension and job type and that various job performance dimensions related to the same trait differently in both magnitude and direction. On the basis of 645 validity estimates

from 84 studies, Tett, Jackson, Rothstein, and Reddon (1994) reported that significant negative trait–performance relations occurred at a rate 28 times higher than that expected by chance. Similarly, in 31% of trait–performance linkages reported in seven meta-analytic studies reviewed by Tett and Christiansen (2007), credibility intervals for the Big Five included  $+ .10$  and  $- .10$ .

In summary, Murphy's (2009) suggestion that "the jury is still out" (p. 462), implying that nonselective trait–performance relations *could* be the order of the day for personality, is, in our view, inaccurate. Rather, we believe the evidence strongly supports the view that personality scale validity varies in both magnitude and direction as a function of content on both sides of the predictor–criterion equation, calling for methods of determining which traits will be important in a given work situation. Importantly, if theory can be effectively applied to the problem of identifying criterion-valid traits a priori, one would be hard pressed to argue against the use of content validation in personnel selection involving personality. We examine this issue next.

### **Can Theory and Content Validation Identify Criterion-Valid Personality Traits?**

Evidence bearing on the extent to which experts can identify job-relevant traits is reported in Tett, Jackson, Rothstein, and Reddon's (1999) meta-analysis, in which personality trait validity coefficients were coded as derived from either exploratory or confirmatory research designs. Exploratory relations were identified as those obtained from "fishing expeditions," in which all trait scales from a given omnibus measure were assessed as potential predictors of job performance. Confirmatory relations, on the other hand, were identified as those obtained for scales selected as especially likely to predict performance based on a priori rationales aided in some cases by job analysis. Consistent with expectations,

confirmatory criterion validity averaged double that for exploratory studies. This strongly supports the value of linking traits to criteria a priori using content-oriented strategies, and, moreover, that experts are able to identify job-relevant traits.

A meta-analysis by Bartram (2005) provides additional support. On the criterion side, performance measures were organized according to the "Great Eight" constructs developed in previous research. Traits were then aligned to criteria by judging the overlap between trait and criterion content domains. Corrected validity in matched cases averaged well above .20, whereas unmatched traits demonstrated virtually no validity. Finally, Robertson and Kinder (1993) meta-analyzed 20 studies that used the *Occupational Personality Questionnaire*. They reported that trait validities varied considerably across criteria and that job experts successfully identified criterion-relevant and criterion-irrelevant traits across 10 job performance criteria.

Another avenue of empirical evidence relevant to content validation in the case of personality tests is that pertaining to the use of narrow versus broad trait scales. In reviewing this literature, Rothstein and Goffin (2006) concluded that, of the studies comparing broad with narrow traits in their relative prediction and incremental prediction, narrow traits always do as well as or better than do broad factors. These findings are relevant to content validation because conceptual mapping of traits to criteria is facilitated by greater articulation (i.e., specification) of the two respective domains (cf. Tett, Guterman, Bleier, & Murphy, 2000). Tett, Steele, and Beauregard (2003), for example, found that the curiosity facet of intellectance on the *Hogan Personality Inventory* correlated positively with technical performance in blue-collar workers, whereas the culture facet correlated negatively. Validity otherwise clouded by the use of broad measures in this case ( $r = \text{near } 0$ ) emerged through reliance on trait facets that were selectively relevant to the nature of the criterion and population.

In summary, personality traits deemed job relevant through the use of content-based analysis tend to show substantially stronger criterion validity than that of traits selected randomly, and reliance on narrow over broad traits facilitates content validation applied to personality tests. We conclude our commentary by considering a theoretical perspective that can help explain *how* and *why* content validity is linked to criterion validity in the case of personality testing.

### **Trait Activation Theory: Untapped Potential for Content Validation in the Case of Personality Tests**

Trait activation theory is based on the well-known concept of person-situation interactionism plus several newer ideas aimed at identifying specific situational features potentially useful in judging the job relevance of a given trait. The main tenet of the theory is that a given trait will be activated—thereby allowing an individual's level on that trait to affect his or her performance—only to the extent that the work setting provides *trait-relevant cues for its expression*. Content validation efforts can be expected to yield greater success (in identifying job-relevant traits) when guided by this and related principles.

Tett and Burnett (2003) discuss how trait activation theory might be applied using content validity methods as a basis for personality-oriented job analysis. At each of the task, social, and organizational levels, a personality trait can be rated for its job relevance using a classification of situational features, including *job demands, distracters, and constraints*. Demands are trait-relevant cues, responses to which are valued positively as job performance. Distracters are opportunities to express traits that hinder performance (e.g., having a sociable manager could result in more slacking for extraverted subordinates). Constraints minimize opportunities to engage in behaviors expressive of a certain trait (e.g., a teleworker who works independently and

autonomously from home has less opportunity to express interpersonal traits such as cooperativeness). According to trait activation theory, a given trait will be judged relevant to the degree the work situation has abundant demands and distracters and where constraints are weak. A positive validity coefficient will be expected in cases where demands outweigh distracters, and a negative coefficient will be expected where the reverse holds.

Demonstrating criterion validity is, by itself, a hollow victory if it is not accompanied by content validation that promotes an *understanding* of why and how the measured trait operates in a given context. Trait activation theory offers a foundation for applying content validation to determine how and why certain traits will be related to job performance in specific jobs. Personality-oriented job analysis methods have yet to fully embrace the situational features outlined in trait activation theory, particularly with respect to demands, distracters, and constraints, as well as the application of these features at the task, social, and organizational levels. Thus, despite the considerable empirical evidence reviewed above, we suggest the full potential of content validation strategies has yet to be realized in the case of personality tests. In addition, the evidence reported above favoring selective trait–performance linkages and the value of confirmatory research suggests criterion validity estimates for personality tests can be expected to improve when driven by appropriate theory.

## Conclusion

Our response is aimed specifically at Murphy's (2009) suggestion that content validity might not matter in personality testing. To the contrary, our review suggests that personality traits may demonstrate very different validities across jobs, that such selective validities are identifiable by experts, and that use of narrow traits and more complete reliance on theory aimed at the trait–situation interface can be

expected to further enhance the usefulness of content validation strategies in the case of personality tests. In light of our analysis, questioning the potential for content validation strategies applied to personality tests seems unwarranted.

## References

- Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*, 1185–1203.
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*, 40–57.
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*, 100–112.
- Hough, L. M. (1992). The "Big Five" personality variables—Construct confusion: Description versus prediction. *Human Performance, 5*, 139–155.
- Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance, 2*, 183–200.
- Murphy, K. R. (2009). Content validation is useful for many things, but validity isn't one of them. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 453–464.
- Murphy, K. R., & Shiarella, H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology, 50*, 823–854.
- Raymark, P. H., Schmit, M. J., & Guion, R. M. (1997). Identifying potentially useful personality constructs for employee selection. *Personnel Psychology, 50*, 723–736.
- Robertson, I. T., & Kinder, A. (1993). Personality and job competencies: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology, 66*, 225–244.
- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16*, 155–180.
- Tett, R. P., & Burnett, D. B. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500–517.
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*, 967–993.
- Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and content validation of a "hyperdimensional" taxonomy of managerial competence. *Human Performance, 13*, 205–251.

- Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick, and Hunter (1994). *Personnel Psychology, 47*, 157–172.
- Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1999). Meta-analysis of bidirectional relations in personality-job performance research. *Human Performance, 12*, 1–29.
- Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior, 24*, 335–356.