Industrial and Organizational Psychology, **2** (2009), 497–500.
Copyright © 2009 Society for Industrial and Organizational Psychology. 1754-9426/09

Clarifying the Concept and Context of Content Validation

BRIAN H. KIM
Occidental College

FREDERICK L. OSWALD
Rice University

Murphy (2009) argues convincingly that although content validation commonly enhances the apparent job relevance and legal defensibility of personnel selection systems, its influence on criterion-related validity is often theoretically and empirically unsupported. In part, this might be because the term *content validation* often reflects a mix of conceptual and legal concerns (American Educational Research Association, 1999; Buster, Roth, & Bobko, 2005; Kleinman & Faley, 1978; Wollack, 1976), and that may obscure those aspects of content validation that improve the test validation process. Murphy's main thesis is that—at its worst—the requirement for establishing a ''manifest relationship'' dictated by the Civil Rights Act of 1964 implies that either (a) the content of relatively abstract selection tests must appear superficially relevant to job performance (e.g., a spatial test for machinists should contain pictures of gears instead of pictures of abstract shapes, even when the underlying questions are the same) or (b) the content of relatively concrete selection tests must match criterion content very closely, if not literally. Murphy correctly claims that neither situation guarantees the high levels of validity (also see Buster et al., 2005; SIOP *Principles*, 2003).

However, we argue that—at its best—content validation is not a narrow prescription for demonstrating superficial linkages between predictor and criterion content. Instead, it represents a broader process using supported theories, past research, and job-related information (e.g., SME data, job analysis, and the O*NET) (a) to determine constructs and construct relationships relevant to personnel selection and (b) to evaluate how the content of selection and criterion measures are faithful to those constructs (Binning & Barrett, 1989). From this perspective, content validation offers a mechanism for improving criterion-related validity by measuring constructs important to selection (i.e., reducing deficiencies) and reducing the measurement of irrelevant factors (i.e., reducing contamination; Hinkin & Tracey, 1999; Little, Lindenberger, & Nesselroade, 1999). A content validation process that leads to incorporating more job-relevant and psychometrically sound predictor and criterion measures into a personnel selection system should add real-world value by reducing actual selection errors (i.e., selecting more of the applicants you want and fewer that you don't). Correlations alone do not provide convincing

Correspondence concerning this article should be addressed to Brian H. Kim.
E-mail: briankim@oxy.edu
Brian H. Kim, Department of Psychology, Occidental College; Frederick L. Oswald, Department of Psychology, Rice University
Department of Psychology, Occidental College, 1600 Campus Rd, F-11, Los Angeles, CA 90041–3314

https://doi.org/10.1111/j.1754-9434.2009.01181.x Published online by Cambridge University Press

validation evidence without a clear rationale underlying the constructs and measures they describe.

Furthermore, two points relevant to Murphy's conclusions deserve further elaboration. First, the usefulness of specific versus general test content clearly depends on the purpose of selection. General cognitive ability (or psychometric *g*) measures typically produce practically significant levels of criterion-related validity for predicting task performance across most jobs, if not all of them (Schmidt & Hunter, 1998). Conversely, a measure saturated with specific content about job tasks could have zero validity if applicants would not have knowledge of such content until posthire (e.g., entry-level sales positions, where product-specific knowledge is learned in training). In support of Murphy's assertions, measuring job knowledge in this case would contribute significantly less to validity than would predictors of knowledge *acquisition*, such as general cognitive ability (particularly fluid intelligence, reflecting logical and spatial relationships) and motivational characteristics (e.g., Deadrick, Bennett, & Russell, 1997; Eyring, Johnson, & Francis, 1993; Kanfer & Ackerman, 1989; Yeo & Neal, 2004).

That said, evidence of positive manifold, or the fact that ability constructs are hierarchically related and are highly correlated (Carroll, 1993), provides no guarantee that specific ability measures are interchangeable predictors of job performance (e.g., substituting a verbal ability measure for a spatial ability measure). If two ability measures correlate .50—the level that Murphy claimed reflected high-positive manifold—and one has a criterion-related validity of .45, then mathematically, criterion-related validity of the other predictor could take on values anywhere from −.55 to +.99 (see Stanley & Wang, 1969), with the largest amount of incremental validity possible at these extremes. Murphy points to a wealth of accumulated evidence for a lack of differential validity across jobs for ability measures, which cannot be ignored, but these results may reflect the lack of precision in administrative ratings of job performance, the complexity of the jobs investigated, and the lack of information on mediating variables (e.g., job knowledge and motivation) as much as the effects of positive manifold, if not more so.

Our second point is an implication of the first. Despite strong validity generalization evidence, the level of criterion-related validity achieved by using general mental ability measures alone often leaves much variance in job performance to be predicted, particularly for complex jobs (Hunter & Hunter, 1984). Complex jobs may require applicants at the outset to possess job-specific knowledge and skills that are highly specialized (e.g., neurosurgeon, industrial–organizational psychologist) and indicative of not only general cognitive ability but also years, if not decades, of motivation invested in a specific knowledge or skill domain (Ackerman, 1996; Cattell, 1971); furthermore, research shows that positive manifold between ability measures tends to be lower among individuals with higher levels of *g* (Detterman & Daniel, 1989; for an example specific to the ASVAB, see Legree, Pifer, & Grafton, 1996). Thus, specific knowledge and skills captured by content validated measures (e.g., knowledge or work-sample tests) may predict variance in performance that more general measures cannot capture (e.g., resumés, license/certification, and interview information) for complex jobs.

More generally, without content validation one may overlook personality, motivation, and interest constructs shown to predict a vast array of performance criteria recognized during the past 15 to 20 years, such as organizational citizenship behavior (e.g., altruism and civic virtue; see Hoffman, Blair, Meriac, & Woehr, 2007) and counterproductive work behavior (e.g., absenteeism and theft; see Gruys & Sackett, 2003). Although ability tests clearly predict whether applicants ''can do'' technical tasks, other measures may be better suited to predict what applicants ''will do,'' and both may be necessary for understanding performance behaviors (see Dilchert, Ones,

https://doi.org/10.1111/j.1754-9434.2009.01181.x Published online by Cambridge University Press

Davis, & Rostow, 2007; Dudley & Cortina, 2008). Personality and biodata measures tend to predict the latter, and the incremental validity of situational judgment tests over ability and noncognitive measures (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004) may in part be the result of asking applicants what they would typically do in specific, job-relevant situations (Motowidlo, Hooper, & Jackson, 2006).

In conclusion, the purpose of our commentary was to share Murphy's skepticism about the usefulness of mechanically linking content between predictor and criterion measures, while also exploring the merits of content validation as a broader process that draws on sources of job-relevant knowledge, theory, and expertise that bolsters substantive explanations about underlying psychological work-related constructs and their relationships. The former is clearly not a substitute for construct or criterion-related validity (Guion, 1977; Landy, 1986), but the latter clearly supplements them. Judgments about appropriate predictor constructs and their content should not be made independent of an understanding of the job or the desired knowledge, skills, abilities, and other characteristics of job incumbents. Validity coefficients for a set of general cognitive measures that are robust and practically significant do not preclude the need to understand how selection for job-specific knowledge and interest will influence specific job functions. We worry that an overreliance on the validity for general cognitive ability may slow the development of other validation efforts to predict a greater portion of the performance domain.

Content validation is a legitimate and valuable process intimately tied to a job analysis of person, task, and situational constructs and characteristics for improving the selection of applicants (Prien, 1977). We (like Guion, 1977, 1978) are sensitive to the fact that experts often disagree, theory is not perfect, and information and interpretations from job analysis and empirical research are not without fault. But ultimately, both expert consensus and points of disagreement arrived at through content validation efforts should be useful for informing both criterion-related validities and validation research, thereby improving the personnel selection process as a whole.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests and knowledge. *Intelligence*, *22*, 227–257.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*, 478–494.

Buster, M. A., Roth, P. L., & Bobko, P. (2005). A process for content validation of education and experience-based minimum qualifications: An approach resulting in federal court approval. *Personnel Psychology*, *58*, 771–799.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies*. New York: Cambridge University Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. North Holland: Amsterdam.

Deadrick, D. L., Bennett, N., & Russell, C. J. (1997). Using hierarchical linear modeling to examine dynamic performance criteria over time. *Journal of Management*, *23*, 745–757.

Detterman, D. K., & Daniel, M. H.. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*, 349–359.

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology*, *92*, 616–627.

Dudley, N. M., & Cortina, J. M. (2008). Knowledge and skills that facilitate the personal support dimension of citizenship. *Journal of Applied Psychology*, *93*, 1249–1270.

Eyring, J. D., Johnson, D. S., & Francis, D. J. (1993). A cross-level units of analysis approach to individual differences in skill acquisition. *Journal of Applied Psychology*, *78*, 805–814.

Gruys, M. L., & Sackett, P. R. (2003). Investigating the dimensionality of counterproductive work behavior. *International Journal of Selection and Assessment*, *11*, 30–42.

Guion, R. M. (1978). ''Content validity'' in moderation. *Personnel Psychology*, *31*, 206–213.

Guion, R. M. (1977). Content validity–the source of my discontent. *Applied Psychological Measurement*, *1*, 1–9.

Hambrick, D. Z., Pink, J. E., Meinz, E. J., Pettibone, J. C., & Oswald, F. L. (2008). The roles of ability, personality, and interests in acquiring current events knowledge: A longitudinal study. *Intelligence*, *36*, 261–278.

https://doi.org/10.1111/j.1754-9434.2009.01181.x Published online by Cambridge University Press

Hinkin, T. R., & Tracey, J. B. (1999). An analysis of variance approach to content validation. *Organizational Research Methods*, *2*, 175–186.

Hoffman, B. J., Blair, C. A., Meriac, J. P., & Woehr, D. J. (2007). Expanding the criterion domain? A quantitative review of the OCB literature. *Journal of Applied Psychology*, *92*, 555–566.

Hunter, J. E., & Hunter, R. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657–690.

Kleiman, L. S., & Faley, R. H. (1978). Assessing content validity: Standards set by the court. *Personnel Psychology*, *31*, 701–713.

Landy, F. J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist*, *41*, 1183–1192.

Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence*, *23*, 45–57.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When ''good'' indicators are bad and ''bad'' indicators are good. *Psychological Methods*, *4*, 192–211.

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, *91*, 749–761.

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.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*, 187–207.

Prien, E. P. (1977). The function of job analysis in content validation. *Personnel Psychology*, *30*, 167–174.

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*, 262–274.

Society for Industrial and Organizational Psychology, Inc. (2003). Principles for the validation and use of personnel selection procedures (4th ed.). Bowling Green, OH: Author.

Stanley, J. C., & Wang, M. D. (1969). Restrictions on the possible values of r_{12} given r_{13} and r_{23}. *Educational and Psychological Measurement*, *29*, 579–581.

Wollack, S. (1976). Content validity: Its legal and psychometric basis. *Public Personnel Management*, *5*, 397–408.

Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology*, *89*, 231–247.

https://doi.org/10.1111/j.1754-9434.2009.01181.x Published online by Cambridge University Press