RESPONSE

Is Content-Related Evidence Useful in Validating Selection Tests?

KEVIN R. MURPHY
Pennsylvania State University

Abstract

The 12 papers commenting on K. R. Murphy (2009a) raise a number of important issues, most of which can be subsumed in one of four themes. First, papers examining content-oriented validation strategies are still necessary and useful, in part because of the frequent use of these strategies in the practice of industrial–organizational (I–O) psychology. Second, the term ''content validity'' means many different things both within and beyond the field of I–O psychology, and it is useful to understand what sorts of inferences examinations of test content do and do not support. Third, these 12 papers present very little evidence that content validation, as typically carried out by I–O psychologists, actually provides information about the likelihood that people who do well on the test will do well on the job. Finally, I believe that the best use of content-related evidence in validating selection tests is in *developing* hypotheses about relationships between test scores and criteria rather than in *testing* these hypotheses.

Is Content-Related Evidence Useful in Validating Selection Tests?

In two recent papers, Murphy and his colleagues have suggested that a procedure that is at the heart of assessments of content validity—that is, evaluations of the match or mismatch between the content of the test and the content of the job—is often irrelevant to determining whether people who do well on those tests will also be likely to perform well on the job (Murphy, 2009a; Murphy, Dzieweczynski, & Yang, 2009). In Murphy (2009a), I took a somewhat strong stand, hoping to stimulate a spirited discussion of content-oriented validation strategies. Comments by Binning and LeBreton (2009), Davison

and Bing (2009), Highhouse (2009), Kim and Oswald (2009), Goldstein and Zedeck (2009), Guion (2009), O'Neill, Goffin, and Tett (2009), Putka, McCloy, Ingerick, O'Shea, and Whetzel (2009), Spengler, Gelléri, and Schuler (2009), Tan (2009), and Thornton (2009), along with a number of ''what the hell?'' e-mails from friends and colleagues, suggests that my paper did indeed stimulate discussion and debate.

These comment papers raised a number of important and useful points. Tan (2009) highlights importance of judgment in both content- and criterion-related studies and reminds us that both methods of validation involve subjective judgment. O'Neill et al. (2009) note that content-oriented validation strategies are likely to have considerable value when applied to personality tests. Murphy et al. (2009) made some similar suggestions, but I believe O'Neill et al. are right that my interpretation of this body of research is probably too conservative and that concerns raised by Murphy (2009a)

Correspondence concerning this article should be addressed to Kevin Murphy.
E-mail: krm10@psu.edu
    Address: Department of Psychology, Pennsylvania State University, Moore Bldg., University Park, PA 16802.

and Murphy et al. are more likely to apply to cognitive domains than to the domain of personality testing. O'Neill et al.'s paper also articulates a number of mechanisms (e.g., trait activation) by which particular personality traits are likely to be linked to at least general types of jobs. In a related vein, Davinson and Bing (2009) discuss evidence supporting the hypothesis that anchoring personality measures in the context in which they will be used, which could be thought of as an extension of content matching, can contribute to criterion-related validity.

Highhouse (2009) makes the important point that content adequacy is not necessarily the same thing as job relatedness; his paper also highlights the sometimes inconsistent treatment of content-oriented strategies in the various editions of the *Standards*. Spengler et al. (2009) note the importance of the criterion (and the frequent inadequacy of criterion measures) for interpreting both content-related and criterion-related evidence. Finally, numerous papers point out the importance of obtaining multiple types of validation evidence (e.g., Binning & LeBreton, 2009; Putka et al., 2009; Thornton, 2009).

In reacting to these comment papers, my first observation was that every methodological paper I have ever written, and most of those I have reviewed, have received two kinds of reviews, one claiming that everyone already knows and understands the point the paper is trying to make and the other claiming that the paper is wrong. These two threads are present in the set of comment papers, with some authors suggesting that there is not much new here (e.g., Goldstein & Zedeck, 2009; Guion, 2009) and others suggesting that my ideas are wrong in a variety of ways (e.g., Binning & LeBreton, 2009; Highhouse, 2009; Thornton, 2009). Despite these differences in perspective, I see a good deal of common ground.

These 12 comment papers cover a range of issues, and it is neither feasible nor productive to write a point-by-point reply. Rather, I think it is useful to comment on four large themes that emerge from a reading of my paper and these 12 commentaries. First, Guion's (2009) question of whether and why a paper like mine is still necessary strikes me as important. Second, this set of papers reinforces my perception that the term ''content validity'' means very different things to different authors. Third, I think it is important to comment on what was *not* included in this set of papers. In particular, I was struck that this set of 12 papers presented little substantial evidence that content validation, as it is typically defined in industrial–organizational (I–O) psychology, actually provides information about the likelihood that people who do well on the test will do well on the job. I hoped that my claim that there is little real evidence to support the most common methods of collecting and interpreting content-oriented validation methods would lead proponents of these methods to bring that evidence forward. A few interesting studies were indeed cited, but I came away from reading these 12 papers more convinced than ever that there is not much evidence that content-oriented validation strategies really work as a proxy for criterion-related validation. Finally, I think it is useful to examine the role of examinations of test content in validation.

Before talking about these four broad themes, it is useful to recap the main point of my critique of content-oriented methods of validation.

What is wrong with content-oriented validation methods? Reading these 12 papers, it is possible to lose sight of exactly what Murphy (2009a) and related papers (e.g., Murphy et al., 2009) say about content validity. My main point in these papers is that if a set of tests that might be used in selection shows positive manifold (i.e., consistent positive correlations, even small ones, with each other and the criterion), the relationship between test scores and measures of performance (i.e., criterion-related validity) does not typically have anything to do with assessments of whether the content

of tests matches or fails to match the content of jobs. That is, content-related validity evidence often turns out to be a poor proxy for evidence about whether or not test scores will predict job performance.

In Murphy (2009a), I used an example of a testing firm that developed content-valid test batteries for entry-level machine operator jobs in one organization and for data entry clerks in another (both contained reliable measures of seemingly job-related knowledge, skills, and abilities [KSAs]). As a result of a mix-up in the mailroom, the test batteries are sent to the wrong organizations. What would happen if each organization went ahead and used the wrong test batteries? Murphy (2009a) and Murphy et al. (2009) cite a good deal of evidence that suggests these batteries will work just about as well in the ''wrong'' job as they would in the job for which they were designed. So much for content matching!

Consider another example. A test publisher develops 20 reliable measures of KSAs that are thought to be relevant to a variety of jobs. Your task is to develop a four-test battery for selecting bus drivers. If you select tests on the basis of a careful job analysis, you will almost certainly end up with a test battery that predicts job performance. However, if you close your eyes and reach into a barrel that contains these 20 tests, pulling four out at random, you will also almost certainly end up with a test battery that also predicts job performance. In fact, you are likely to achieve similar levels of criterion-related validity with both test batteries (Murphy, et al.). If you select tests solely on the basis of their relationship with the general cognitive factor (i.e., their *g* loading), you will probably do an even better job predicting performance (Ree & Earles, 1991; Ree, Earles, & Teachout, 1994). Again, so much for content matching!

My critique of content-oriented validation methods is that they rely on evidence that turns out to have very little to do with the validity of tests as predictors of performance. Assessments of the match or the linkage between test content and job content, which are at the heart of all content-oriented methods of validation, certainly *seem* like they should be relevant to the validity of tests as predictors of performance, but often they are *not* relevant (Murphy et al., 2009). In contrast, if the issue is whether or not test scores will predict future performance, and the tests in question show positive manifold, content validity has little to do with criterion-related validity. Later in this paper, I will offer some suggestions for the more fruitful use of content-oriented investigations when validating tests.

Was this trip necessary? Guion (2009) poses the question of whether papers like Murphy (2009a) are necessary, and he concludes (sadly, I think) that it still is necessary to debate and discuss issues psychometricians regarded as largely settled decades ago. I think this trip is *very* necessary because the practice of I–O psychology still relies so heavily on content-oriented validation strategies.

The field of I–O psychology has long prided itself on its successful marriage of science and practice (Murphy & Saal, 1990), but I am not always certain we should be so smug. My own experiences as an Editor, a consultant, an expert witness, and a workshop presenter suggest that there are many well-trained I–O psychologists who routinely use content validation (in particular, expert judgments of the overlap between tests and jobs) as the primary or the sole class of evidence for the validity of tests as predictors of job performance. Tan (2009) articulated what I believe is a common view among I–O practitioners,[1] noting ''Conducting criterion-related validation studies are often not feasible because of the time and expense associated with it.... It is often much quicker and less expensive to use a content validation strategy when developing a test.''

It is hard to know with any certainty what validation strategies are most often applied in different settings, and I am sure there are many researchers, consultants, and firms who routinely use a range of sophisticated approaches when validating tests. Nevertheless, I have seen many validity reports that are based entirely on content-oriented strategies, which leads me to think that examinations of content-oriented methods of validation still have important implications for understanding the links between science and practice in I–O psychology. The Guions, Cronbachs, and other eminent measurement specialists of the world have been telling us for over 30 years that ''… there is no such thing as content validity'' (Guion, 1978a p. 212) *but we have not been listening*. Goldstein and Zedeck (2009) refer to my angst over content validity. If I do feel angst, it is a result of the apparent disconnect between science and practice in many of the validity studies I have seen.

How did we get in this mess? One likely explanation is that the practice of I–O psychology is strongly affected by both scientific and legal considerations (Kim & Oswald, 2009; Tonowski, 2009). As a result, the terms we use and the strategies we pursue in validating tests are not completely our own; practitioners must balance the sometimes competing demands of science, the law, and their clients in determining whether or not to recommend particular tests. It is little surprise that the concerns of psychometricians have not trumped the concerns of other stakeholders, especially when content-oriented validation strategies *seem* so logical and effective.

What does ''content validity'' mean? A review of professional guidelines, legal requirements, established practices in I–O psychology, and published literature dealing with content-oriented strategies for validating selection tests suggests the term ''content validity'' could refer to any of at least three methods for evaluating or demonstrating the validity of tests and

assessments as predictors of future performance on the basis of assessments of the match between test content and job content. The three approaches described below (domain representativeness, performance requirements, and linkage approach) do not necessarily exhaust the realm of possibilities, but they do encompass the different approaches to content-oriented validation I have encountered in the literature, in litigation, in consulting, and in discussions with colleagues in the field of I–O psychology.

Domain representativeness. One approach laid out in the *Uniform Guidelines on Employee Selection Procedures* (1978) is to argue that the test is a valid predictor of performance if it can be shown to be a representative sample of the work domain. Section 14 C(4) of the *Guidelines* states: ''To demonstrate the content validity of a selection procedure, a user should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job.''

Similarly, the third edition of the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, Inc., 1987) stated that a test is content valid if it is ''… a representative sample of the tasks, behaviors, or knowledge drawn from that [job] domain'' (p. 19). Guion (1998) noted that tests are not required, under this definition, to sample the entire job, but that the critical job functions must be represented in the test. Under this definition, content validity can be established if experts agree that a test adequately samples the portions of the job domain that are critical to adequate performance. This strategy is based on the inference that people who do well on a sample of the job will also do well if selected into that job.

The domain representativeness definition of ''content validity'' is probably the most dated and the least common approach in current validation research, but this

method is sometimes still used to demonstrate the validity of work sample tests and other assessment methods that are designed to closely mimic the behavior the tests are designed to predict.

Performance requirements approach. An alternate strategy is to describe a test battery as content valid if it measures knowledge, abilities, and/or skills that are required for successful job performance. Section 14 C(1) of the *Uniform Guidelines* (1978) notes: ''Selection procedures which purport to measure knowledge, skills, or abilities may in certain circumstances be justified by content validity, although they may not be representative samples, if the knowledge, skill, or ability measured by the selection procedure can be operationally defined . . ., and if that knowledge, skill, or ability is a necessary prerequisite to successful job performance.''

Similarly, the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, Inc., 2003) state ''Evidence based on test content may include logical or empirical analyses that compare the adequacy of the match between test content and work content, worker requirements, or outcomes of the job'' (p. 454), and ''Evidence for validity based on content typically consists of a demonstration of a strong linkage between the content of the selection procedure and important work behaviors, activities, worker requirements, or outcomes on the job'' (p. 462).

Cascio (1998) provides a detailed description of the performance requirements approach. As Cascio describes this method, content-related validity can be evaluated in terms of the extent to which subject matter experts (SMEs) perceive overlap between the KSAs the test measures and the KSAs essential for job performance. This process is typically carried out in several steps: (a) each SME is given a set of test items and independently indicates whether the knowledge or skill measured by the item is essential, useful but not essential, or not necessary to the performance of the job, (b) responses from all SMEs are pooled and

the number indicating ''essential'' is determined, (c) a content validity ratio (CVR) is determined for each test item [CVR = $(n_e - N/2)/(N/2)$, where n_e is the number of SMEs indicating ''essential'' and N the total number of SMEs], (d) items are eliminated if the CVR fails to meet statistical significance (which can be determined from a table presented in Lawshe, 1975), (e) the mean CVR value of the retained items (the content validity index [CVI]) is then computed. The CVI represents a quantitative estimate of the extent to which perceived overlap exists between capability to function in a job performance domain and performance on the test under investigation.

These steps need not be followed slavishly; the key to this approach to content validation is to establish overlap between what the test measures and what is judged to be essential to job performance (Muchinsky, 1999). The use of quantitative measures, such as CVR and CVI, has some advantages, but it is important to keep in mind that this method relies on judgments, not on empirical evidence of test-performance links.

Linkage methods. Linkage methods represent a variation on the performance requirements approach described above, in which SMEs are asked to make judgments about overlaps in the KSAs required to do well on a test and those required to do well on a job. Goldstein and Zedeck (1996) argue that ''. . . when there is congruence between the KSAs required to perform on the job and the KSA required to perform on the testing instrument, then it should be possible to make inferences about how the test scores relate to job behavior'' (p. 28; see also Goldstein, Zedeck, & Schneider, 1993). This method involves (a) linking KSAs to specific job elements (i.e., making a judgment that specific KSAs are required for or contribute to performance of specific aspects of a job), (b) linking KSAs to test items or to subtests (these steps are often done by independent groups of experts; Goldstein & Zedeck, 1996, refer to this step as retranslation), and (c) assessing the communalities between these KSA lists. If

the same KSAs are judged to be required for performing well on the test and for performing well on the job, then the inference is made that people who do well on the test will also do well on the job.

Linkage methods represent the most sophisticated of the common approaches to content validation. The potential weakness of this method is its exclusive reliance on judgments—in particular judgments about the KSAs that are required for successful job performance. Burns (1996) cautions that claims by SMEs that particular KSAs are required by the job are difficult to verify or even evaluate. Harvey (1991) notes that there can be a large inferential leap between the observable information in a job analysis and judgments about ability requirements or requirements for other constructs. There is evidence that experts are able to make reliable and consistent judgments regarding these job requirements and that empirical validations of these judgments are feasible (Fleishman & Mumford, 1991; Levine, May, Ulm, & Gordon, 1997). However, empirical validation of expert judgments regarding KSA requirements are rarely included in content validity studies.

Common threads. The three strategies described above differ in many specifics, but they all share two important common threads that serve to define content-oriented validation methods. First, all three approaches suggest that validity can be defined in terms of some sort of overlap between the content of the test and the content of the job. This overlap might be defined in terms of similarities between behaviors on the test and behaviors on the job (e.g., domain representation methods) or between the KSAs measured by the test (or those required for successful test performance), but in all cases, it is this judged overlap that supports the inference that test scores are valid predictors of performance. Second, content-oriented validation methods are most likely to be applied to tests designed or used to measure KSAs thought to be relevant to job performance. To be sure, content-oriented methods are not necessarily limited

to measures of job-relevant KSAs, but the application of content-oriented validity strategies as the primary or the sole strategy for determining the validity of tests as predictors of performance appears to be more common in the cognitive domains that encompass most measures of KSAs than in other domains such as personality.

Binning and LeBreton (2009), Highhouse (2009), Kim and Oswald (2009), and Spengler et al. (2009) describe a number of sophisticated ways of using content-related evidence in validation, and as I note in the final section of this paper, there are indeed many ways to take advantage of this sort of evidence. However, I do not believe these sophisticated models describe the reality of practice in I–O psychology. Furthermore, I am not convinced that much of the rigmarole that constitutes a typical content validity study (e.g., CVRs and linkage judgments) has any real value for the purpose of evaluating the likelihood that people who do well on selection tests will or will not tend to do well on the job.

The case of the missing evidence. When I was a graduate student, a series of papers by Guion (1977, 1978a, 1978b) questioning content-oriented validation strategies were required reading. Several years ago, when I revisited these papers while teaching a graduate seminar, I decided to look at the evidence that content-oriented validation strategies actually tell you something about the validity of a test as a predictor of performance. This quickly turned into an exercise of looking *for* the evidence that tests that are judged to be content valid are better predictors of performance on the job than tests that are not judged to be content valid. I did not find much supporting evidence and found a good deal of evidence that was not consistent with the claim that comparisons of test content and job content provide useful information about the likely criterion-related validity of a test.

Two of the papers that commented on Murphy (2009a) discussed sources

of evidence that might support content-oriented validation methods. Tan (2009) discusses a few studies that provide at least indirect evidence that content matching can matter. Spengler et al. (2009) present evidence for the incremental validity of hybrid tests that include a mix of general and specific features (including sections based specifically on job content) over more general tests. However, none of the 12 papers commenting on Murphy (2009a) presents convincing or substantial evidence to support a practice that is widespread in the I–O community—that is, using judgments about job-test content matches to evaluate or demonstrate the validity of tests as predictors of job performance.

Putka et al. (2009) remind us that the absence of evidence supporting content-oriented validation methods is not the same as evidence that these methods do not work. This is true, but the burden of persuasion should be on those I–O psychologists who wish to use content-oriented approaches as the primary or sole strategy for validating tests. There is a good deal of evidence that content-oriented validation strategies do not tell us much about whether or not test scores predict job performance, and not much real evidence that these strategies really work the way many I–O psychologists think they do. I think it is time that the burden of proof should shift to people who continue to use content-oriented strategies as a proxy for criterion-oriented validation. If a responsible psychologist wants to use content validation as the strategy for demonstrating that his or her tests do indeed predict performance, it is time for him or her to bring forth evidence that this approach to validation actually provides evidence that is relevant to the validity of a test as a predictor of performance.

Should we ignore content? Thornton (2009) rightly notes that the *Standards* embrace the use of content-related evidence in validating selection tests. Similarly, Tonowski (2009) notes that the *Guidelines* similarly embrace these procedures. On the other hand, Murphy (2009a) and Murphy et al.

(2009) suggest that these procedures do not always provide useful evidence for the validity of tests as predictors of performance. This does not mean, however, that content-oriented procedures have no place in validation.

First, the term "content validity" often has a different meaning in the rest of the psychometric community than in I–O community. When the purpose of a validity study is to determine whether or not a particular test measures a particular content domain or construct, examinations of test content are very useful. For example, Murphy and Davidshofer (2005) provide detailed examples of content-oriented strategies for evaluating the validity of measurement—that is, the degree to which tests measure what they purport to measure. Outside of the field of I–O psychology, content-oriented validation strategies are a common part of determining what tests measure, but they are rarely used as a proxy for criterion-related strategies. The use of content-oriented validation strategies to determine whether test scores predict performance, effectiveness, or other criteria is almost unique to our field, and is perhaps a legacy of our entanglement with the legal system.

Over the last 30–40 years, there has been a clear trend in thinking about validity, moving from obsessing about different "types" of validity to understanding validity as a unified concept (Landy, 1987; Murphy, 2009a). An unfortunate downside of this trend toward the unification of validity is that it is sometimes easy to lose track of exactly what is being validated and why. For example, suppose an organization uses a measure that appears to tap mechanical ability in its selection process. Validity studies might be used to accomplish two potentially distinct goals—that is, determining whether this test really measures that construct and determining whether scores on this test predict some other distal construct, such as performance or effectiveness. Different strategies and different types of validation evidence might be relevant to these two goals (Murphy & Davidshofer, 2005).

In the context of personnel testing, where the ultimate goal of many validity studies is to determine whether people who do well on tests are also likely to perform well on the job, examinations of test content are more useful for *generating* hypotheses than they are for *testing* them. For example, examinations of test content can provide a basis for developing hypotheses about what construct or constructs the test measures. If a test is interpreted as a sample from a particular content domain, a detailed examination of test content in relationship to the structure and contents of that domain can be an invaluable aid to determining the validity of measurement (Murphy & Davidshofer, 2005).

As our understanding of validation has evolved, it has become clear that all of the different strategies for collecting validity serve the same end—that is, they help you understand the meaning and implications of test scores (Angoff, 1988; Langenfeld & Crocker, 1994; *Standards for Educational and Psychological Testing*, 1999). It has also become clear that it does not make sense to label a test ''valid'' or ''not valid''; validity is a property of the inferences made on the basis of test scores, not a property of the tests themselves (e.g., Lawshe, 1985). Content-related evidence is extremely useful for assessing inferences about what tests measure. These might, in turn, help draw a chain of inference about what other variables these tests scores might predict (e.g., criteria). For example, if the content of a test supports the inference that the test measures a particular construct, *and* a credible body of research on that construct shows that good measures of that construct predict important criteria, a chain of inferences from the test to the criteria is established. However, it is important to understand that the inferential leap is not directly from test content to criteria. Rather, test content helps to establish what a test measures, and research and theory regarding that construct or content domain provide the crucial link from tests to constructs to criteria.

Suppose you design or interpret a test as a work sample. Here, the traditional methods of content validation might seem more appropriate, but even here a stronger and sounder inferential chain can be developed by attempting to identify the constructs measured by the test and taking advantage of the substantial body of research that is likely to be available about those constructs and their relationships to important criteria. A content-oriented strategy answers the question of whether the work sample is indeed a sample of the work behavior you are trying to predict, but even if adequate sampling can be established, there are still a number of differences between the test and the behaviors to be predicted which might diminish the link between scores on work samples and measures of subsequent work behavior. Most obviously, applicants completing a work sample are likely to show maximal performance, which may or may not be a good indicator of their typical performance levels (Sackett, Zedeck, & Fogli, 1988). More generally, testing is carried out in a setting where every effort is made to remove extraneous influences and distractions, whereas work behavior is carried out in a setting where there are many things other than the ability, skill, or competency of the individual that may influence their actual on-the-job performance.

In validation, it is almost always better to collect multiple types of evidence than to rely entirely on one category of evidence to support the inferences that are made on the basis of test scores (Murphy, 2009b; Thornton, 2009). Content-oriented methods have their place in the toolbox of I–O psychologists, but we need to be careful and realistic about what sorts of inferences these methods do and do not support. A careful examination of test content can be very useful for determining what a test measures and for *developing* hypotheses about what criteria test scores will predict. When tests show positive manifold, a comparison of test content and job content does not provide relevant information for *testing*

hypotheses about whether people who do well on the test will also do well on the job.

References

Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 19–32). Hillsdale, NJ: Erlbaum.

Binning, J. F., & LeBreton, J. M. (2009). Coherent conceptualization is useful for many things, and understanding validity is one of them. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 486–492.

Burns, W. C. (1996). Content validity, face validity, and quantitative face validity. In R. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 38–46). Westport, CT: Quorum Books.

Cascio, W. F. (1998). *Applied psychology in human resource management* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Davison, H. K., & Bing, M. N. (2009). Content validity does matter for the criterion-related validity of personality tests. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 500–502.

Fleishman, E. A., & Mumford, M. D. (1991). Evaluating classifications of job behavior: A construct validation of ability requirement scales. *Personnel Psychology*, 44, 523–575.

Goldstein, I. L., & Zedeck, S. (1996). Content validation. In R. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 27–37). Westport, CT: Quorum Books.

Goldstein, I. L., & Zedeck, S. (2009). Content validity and Murphy's angst. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 495–495.

Goldstein, I. L., Zedeck, S., & Schneider, B. (1993). An exploration of the job analysis-content validity process. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 3–34). San Francisco: Jossey-Bass.

Guion, R. M. (1977). Content validity–The source of my discontent. *Applied Psychological Measurement*, 1, 1–10.

Guion, R. M. (1978a). ''Content validity'' in moderation. *Personal Psychology*, 31, 205–213.

Guion, R. M. (1978b). Scoring content domain samples: The problem of fairness. *Journal of Applied Psychology*, 63, 499–506.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel selection*. Mahwah, NJ: Erlbaum.

Guion, R. M. (2009). Was this trip necessary? *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 465–468.

Harvey, R. J. (1991). Job analysis. In M. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 71–164). Palo Alto, CA: Consulting Psychologists Press.

Highhouse, S. (2009). Tests don't measure jobs: The meaning of content validation. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 493–495.

Kim, B. H., & Oswald, F. L. (2009). Clarifying the concept and context of content validation. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 496–499.

Landy, F. J. (1987). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist*, 41, 1183–1192.

Langenfeld, T. E., & Crocker, L. M. (1994). The evolution of validity theory: Public school testing, the courts, and incompatible interpretations. *Educational Assessment*, 2, 149–165.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563–575.

Lawshe, C. L. (1985). Inferences from personnel tests and their validities. *Journal of Applied Psychology*, 70, 237–238.

Levine, E. L., May, D. M., Ulm, R. A., & Gordon, T. R. (1997). A methodology for developing and validating minimum qualifications (MQs). *Personnel Psychology*, 50, 1005–1023.

Muchinsky, P. M. (1999). *Psychology applied to work* (6th ed.). Belmont, CA: Wadsworth.

Murphy, K. R. (2009a). 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. (2009b). Validity, validation and values. *The Academy of Management Annals*, 3, 421–461.

Murphy, K., & Davidshofer, C. (2005). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Murphy, K. R., Dzieweczynski, J. L., & Yang, Z. (2009). Positive manifold limits the relevance of content-matching strategies for validating selection test batteries. *Journal of Applied Psychology*, 94, 1018–1031.

Murphy, K., & Saal, F. E. (1990). *Psychology in organizations: Integrating science and practice*. Hillsdale, NJ: Erlbaum.

O'Neill, T. A., Goffin, R. D., & Tett, R. P. (2009). Content validation is fundamental for optimizing the criterion validity of personality tests. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 508–512.

Putka, D. J., McCloy, R. A., Ingerick, M., O'Shea, P. G., & Whetzel, D. L. (2009). Links among bases of validation evidence: Absence of empirical evidence is not evidence of absence. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 475–480.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. *Personnel Psychology*, 44, 321–332.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology*, 79, 518–524.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relationship between measures of typical and maximal job performance. *Journal of Applied Psychology*, 73, 482–486.

Society for Industrial and Organizational Psychology, Inc. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park, MD: Author.

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

Spengler, M., Golléri, P., & Schuler, H. (2009). The construct behind content validity: New

approaches to a better understanding. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 503–507.

Standards for Educational and Psychological Testing. (1999). Washington, DC: American Psychological Association.

Tan, J. A. (2009). Babies, bathwater, and validity: Content validity is useful in the validation process. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 513–515.

Thornton, G. C. (2009). Evidence of content matching is evidence of validity. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 469–474.

Tonowski, R. F. (2009). ''Content'' still belongs with ''validity.'' *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 481–485.

Uniform Guidelines on Employee Selection Procedures. (1978). 29 C.F.R. 1607.