Industrial and Organizational Psychology, **3** (2010), 356–359.
Copyright © 2010 Society for Industrial and Organizational Psychology. 1754-9426/10

Synthetic Validity: A Great Idea Whose Time Never Came

KEVIN R. MURPHY
The Pennsylvania State University

Synthetic validity is very much like monogamy. It is an idea that is universally admired but not widely practiced. The concept of synthetic validity has been around for almost 60 years (Lawshe, 1952) and it is described in virtually every testing and measurement book and every industrial and organizational (I-O) psychology textbook I have ever read. I still vividly remember my undergraduate Psychological Testing professor dismissing synthetic validity as ''made-up validity.'' I thought he was wrong then, and I still do. Nevertheless, it is worth asking why synthetic validity studies are so rare.

There is no doubt that synthetic validity is a great idea. Johnson et al. (2010) do a masterful job laying out the history of synthetic validity, its major assumptions and components, and the many advantages of synthetic validity. They are also realistic (but optimistic) about meeting the challenges of synthetic validity. They argue convincingly that synthetic validity could be a key to advancing both the science and the practice of personnel selection. Indeed, my first reaction on reading their article was ''what's not to like about synthetic validity?'' Given all the good ideas built into synthetic validity, how can we explain its failure to catch fire?

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

There is little doubt that synthetic validity has had limited application over the last 50 years (Guion, 2006; Murphy, 2009b). I believe that the failure of this approach to truly catch on is partly a matter of bad luck. By the time the methods were developed that allowed I-O psychologists to address some of the challenges of conducting synthetic validity studies (e.g., validity generalization, dependable databases describing jobs, and validity study outcomes), the need for synthetic validity was no longer so clear. It is worth remembering that Lawshe (1952) introduced this concept in the context of the problems faced by small organizations, or in validity studies where the total number of incumbents is small. That is, synthetic validity was introduced largely in the context of something you could do if practical considerations (especially small samples) made a local validity study impossible. As Johnson et al. note, this concern with small samples and practicality might be phrased in more general terms, allowing us to think about synthetic validation as a system for drawing inferences about the validity of various selection tests and test batteries in the absence of much local information other than a reasonably general description of the job. Finally, synthetic validity could be thought of as a system for answering the question, ''how can we design a valid selection test battery on the basis of general scientific principles and knowledge rather than doing so on the basis of local knowledge?''. Unfortunately, the need for a system like synthetic validity for answering

https://doi.org/10.1111/j.1754-9434.2010.01251.x Published online by Cambridge University Press

these questions is no longer there. That ship has sailed.

First, the question is no longer whether one should do local validity studies when samples are small, but rather whether one should do local validity studies at all (Murphy, 2009b; Schmidt & Hunter, 1998). Proponents of validity generalization might argue that we have accumulated enough knowledge about the validity of several classes of selection tests, particularly those that measure general cognitive ability, that new local validity studies can rarely if ever produce useful knowledge. If they are done in small samples, local validity studies produce results that are inherently untrustworthy (e.g., with a sample of 100, a 95% confidence interval around an observed correlation coefficient will be approximately .20 points wide in either direction). If you put forward synthetic validity as an alternative to doing small-sample, local validity studies, one answer is that simpler meta-analytic alternatives already exist.

Second, synthetic validity might be put forth as an answer to the question ''how should we design selection tests so that we can be reasonably certain that people who do well on the test will do well on the job?''. I think, at least in a broad sense, we already know the answer to this question. In just about any job in just about any organization, good measures of general cognitive ability, perhaps together with good measures of broad personality traits such as Conscientiousness, will do a pretty good job in making valid predictions about future job performance (Schmidt & Hunter, 1998; Tett, Jackson, & Rothstein, 1991). It is hard to think of any job in which it would not be better to hire applicants who are smart and conscientious than it would be to hire applicants who are stupid and undependable. To be sure, you might do better using tests that are selected specifically on the basis of the content of the job, but you often will not do *much* better (Ree, Carretta, & Earles, 1998; Ree & Earles, 1991, 1992). There are many good reasons for tailoring the content of tests to the content of jobs (e.g., acceptability to applicants and organizations, legal guidelines), but validity is not one of them (Murphy, 2009a). We know a good deal about how to make valid selection decisions in most jobs, and a careful analysis of job demands and tests that are optimal for predicting particular components of the job will often at best make a small contribution to increasing the validity of selection system.

Solving Scientific Versus Operational Problems

Although the practical case for synthetic validity is not as compelling as it once seemed, the scientific case for this approach might still be enough to motivate its further use. To understand some of the uses of synthetic validity in advancing the science of personnel selection, it is useful to think concretely about what it takes to implement synthetic validity. As Johnson et al. note, there are at least two different approaches to synthetic validity (job component validity and job requirements approach), but each starts with a matrix that contains information about the relationships between j general components of jobs and k predictor constructs. This information is used to draw inferences about the validity of a particular set of predictors in a particular job.

To implement synthetic validity, it is necessary to first populate this $j \times k$ matrix relating predictor constructs and job components.

As Murphy and Shiarella (1997) point out, to understand the validity of test batteries in different settings it is also necessary to think about two other matrices: the $j \times j$ matrix of relationships among job components and the $k \times k$ matrix of relationships among predictor constructs. If we think of a job as representing a combination of particular components, the validity of a battery of predictor measures for predicting overall performance in any particular job is determined by the products of matrices of relationships among job components, relationships among predictor constructs,

https://doi.org/10.1111/j.1754-9434.2010.01251.x Published online by Cambridge University Press

and relationships among predictors and components.

I think the compelling scientific questions that arise as a result of applying the synthetic validity model involve determining *how many* and *which* components and predictor constructs should be considered in populating the $j \times k$ matrix of relationships between predictor constructs and job components. Because so many of the predictor constructs we use in personnel selection (particularly those that involve general cognitive ability in some way) are positively intercorrelated (for discussions of positive manifold in predictors, see Murphy, 2009a, 2009b; Murphy, Dzieweczynski, & Yang, 2009), the optimal number of rows in a $j \times k$ matrix of relationships between predictor constructs and job components is probably smaller than what most personnel selection specialists think. That is, it is possible to name many different knowledge, skill, and ability dimensions (as well as other attributes) that *might* form rows in this matrix, but the pervasive pattern of positive intercorrelations among these measures has clear implications for validity. If you attempt to use a number of specific abilities, skills, knowledge, and so forth to form prediction equations for different job, what you will find is that these equations often fail to demonstrate differential validity. As Johnson et al. note, synthetic validity studies do not have a very encouraging track record in terms of the differential validity of prediction equations developed for jobs that differ substantially in content; predictor composites developed specifically for one job or job family turn out to work just about as well in other jobs of similar complexity, even if their content is very different. Murphy et al. demonstrated that this lack of differential validity is an unavoidable consequence of the positive manifold shown by most predictors used in personnel selection. One implication for the strategy of synthetic validity is that predictors that show minimum overlap should probably be chosen, and this will naturally lead to a reduction in both the number and the specificity of the predictor constructs that are chosen to define the rows of the predictor construct by job component matrix.

Similarly, it will be important to determine how to best define the columns of this matrix. One possibility would be to use the O*NET Generalized Work Activities to define the columns of this matrix. Depending on the level of specificity chosen, this could lead to a matrix with more than 40 columns, but it is unlikely that this many columns will be needed. The latent structure of the universe of jobs in today's economy is not completely understood. It might be as simple as data–people–things, or it might be considerably more complex.

Determining the optimal rows and columns to be included in a predictor construct by jobs matrix strikes me as a very important scientific problem, but one that might have a relatively small practical payoff. Even if we can consistently improve on the default advice of basing selection decisions on some combination of cognitive ability and Conscientiousness, I worry that we will still face the problem of developing personnel selection systems that are accepted by applicants, organizations, and regulatory agencies as valid and reasonable, and I think this is something that will continue to require the sorts of careful attention to job analysis, input from key stakeholders, empirical analyses using local criteria, and so forth that characterize more traditional approaches to developing personnel selection systems.

References

Guion, R. M. (2006). Still learning. *The Industrial-Organizational Psychologist, 44*, 83–86.

Johnson, J. W., Steel, P., Scherbaum, C. A., Hoffman, C. C., Jeanneret, P. R., & Foster, J. (2010). Validation is like motor oil: Synthetic is better. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*, 305–328.

Lawshe, C. H. (1952). Employee selection. *Personnel Psychology, 6*, 31–34.

Murphy, K., & Shiarella, A. (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.

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.

https://doi.org/10.1111/j.1754-9434.2010.01251.x Published online by Cambridge University Press

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

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.

Ree, M. J., Carretta, T. R., & Earles, J. A. (1998). In top-down decisions, weighting variables does not matter. A consequence of Wilks' Theorem. *Organizational Research Methods*, *1*, 407–420.

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. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, *1*, 86–89.

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.

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

https://doi.org/10.1111/j.1754-9434.2010.01251.x Published online by Cambridge University Press