Industrial and Organizational Psychology, **3** (2010), 366–370.
Copyright © 2010 Society for Industrial and Organizational Psychology. 1754-9426/10

Developing an Online Synthetic Validation Tool

RODNEY A. McCLOY AND DAN J. PUTKA
Human Resources Research Organization

ROBERT E. GIBBY
Procter & Gamble

In their comprehensive apologetic treatment of synthetic validity, Johnson et al. (2010) echo Hough (2001), advocating development of a central synthetic validation database, which would serve as a repository of validity information to support future synthetic validation efforts. They offer two potential approaches for developing such a database. The first entails the conduct of ''a large-scale study in which tests are administered to and performance ratings gathered on incumbents in a large number of jobs in a variety of organizations.'' The authors consider this approach to be ''ideal but impractical,'' largely because of the scope and cost of the data collection required and the resulting investment required by any sponsoring organization. The second entails the conduct of multiple local studies to generate empirical estimates of relationships between measures of various predictor constructs and a standardized set of job components. Johnson et al. consider this approach more practical, citing Meyer, Dalal, and Bonaccio (2009) as a benchmark example.

Correspondence concerning this article should be addressed to Rodney A. McCloy.
E-mail: rmccloy@humrro.org
Address: Human Resources Research Organization (HumRRO), 10503 Timberwood Circle, Suite 101, Louisville, KY 40223
Rodney A. McCloy and Dan J. Putka, Human Resources Research Organization (HumRRO); Robert E. Gibby, Procter & Gamble.

We propose a third approach for attaining a synthetic validation database that can offer practitioners a valuable resource almost immediately. We argue that perhaps Johnson et al. have underplayed something that resides in their own article: capitalizing on naturally occurring, distributed teams of subject matter experts (SMEs) to generate estimated relationships among variables (e.g., locally used predictors and standard job components) required to synthesize validity estimates for practitioners' local selection situations. Rational judgment approaches to synthesizing validity have demonstrated value in specific selection situations (Peterson, Wise, Arabian, & Hoffman, 2001) and, we argue, could be capitalized on at a more aggregate level by offering practitioners an online tool that provides standardized stimulus materials and data collection procedures based on the Occupational Information Network (O*NET) framework.

Specifically, we envision an Online Synthetic Validity Estimator (OSVE) to which individual practitioners could go to enter basic information gathered via a set of common language stimulus materials for collecting synthetic validity information and, in return for their efforts, obtain synthetic validity estimates tailored to their specific occupations and sets of predictors. Over time, data contributed by practitioners would accrue in the

https://doi.org/10.1111/j.1754-9434.2010.01254.x Published online by Cambridge University Press

database underlying the tool, allowing the body of industrial-organizational (I-O) researchers to examine trends and evaluate potential moderators of local predictor–job component and component–component relationships across the work contexts reflected in the estimates.

A Fledgling Proposal

As a standardized job analysis tool providing data on all occupations in the U.S. economy, O*NET provides a reasonable foundation for a centralized OSVE. Further, rather than incurring the expense of administering predictors and rating scales to incumbents in hundreds of occupations nationwide, we propose to use locally available SMEs, thereby distributing the burden for populating the database across the body of practitioners desiring to obtain synthetic validity estimates for their local selection situations. In concept, the core elements of our approach are similar to those evaluated and described by Peterson et al. (2001), but in form and ultimate function, our approach is more novel, utilizing advances in technology and capitalizing on the collective resources of practitioners in the field.

*Exploiting the O*NET Framework*

The O*NET database provides prospective users with importance ratings for 42 O*NET Generalized Work Activities (GWAs; job components) for the 800+ occupations in the Standard Occupational Classification system. The GWA taxonomy, as well as the importance ratings that populate it, can provide a common framework for helping practitioners gather SME judgments regarding correlations among individuals' performance on GWAs critical to the practitioners' occupations of interest. Having a set of job components expressed in a common language across occupations, coupled with a standard SME judgment task by which estimates are gathered, will allow future users of the database to evaluate the extent to which such ratings (a) are reliable (examine consistency between ratings inputted by different practitioners examining the same job, similar jobs) and (b) depend on the job in question (i.e., whether relationships between components are comparable across jobs or vary in predictable ways as a function of work context). Without a common language and a rigorous, standardized data collection process, such questions would be difficult to assess.

Obtaining Local SME Judgments via Users of the System

As noted, judgments provided by SMEs (identified by local practitioners) would constitute the estimates entered into the system. Depending on the types of data gathered, these SMEs might be I-O psychologists, job incumbents, or supervisors. Essentially, the SMEs would reflect the set of individuals in the best position to estimate the quantities of interest for a practitioner's local situation. Similar to the process used in the U.S. Army's synthetic validation project (Peterson et al., 2001), SMEs would provide estimates of the expected validity for the predictor measures for predicting the relevant job components (GWAs) for the occupations in question, as well as correlations among the predictors and job components. If a practitioner's target occupation matched one already in the O*NET database, GWA importance ratings would serve to filter the 42 GWAs so that SMEs would need to provide estimates for only those GWAs deemed important for the occupations of interest.

Offering Stimulus Materials That Help Ensure Quality Data

The stimulus materials offered by the OSVE would be another key to its success and would include elements such as (a) standard lists of job components and their definitions (from O*NET), (b) instructions for gathering estimates needed as input to the OSVE, (c) minimum requirements and desired characteristics for participating SMEs, (d) scripts for training SMEs from whom estimates would be gathered, and (e) a standardized process for checking and

https://doi.org/10.1111/j.1754-9434.2010.01254.x Published online by Cambridge University Press

ensuring the quality of the resulting data. The stimulus materials would be designed to ensure that practitioners are entering estimates using data that were collected via a rigorous, standardized process that helps ensure their quality and avoids creating a system that suffers from the ''garbage in–garbage out'' criticism sometimes leveled against meta-analysis (cf. Oswald & McCloy, 2003).

Calculating the Synthetic Validity Coefficient

With the data described above, the Nunnally and Bernstein (1994) equation for calculating the correlation between composites would be used to generate a synthetic validity estimate for the given predictor (or set of predictors) of interest and the performance composite comprising the selected GWAs. The online tool would provide this estimate to the practitioner automatically, using the data they had entered. Options for obtaining estimates are described in further detail in the section that follows.

How Would an Online Synthetic Validity Estimator Work?

To demonstrate how an OSVE might work, we present three scenarios that we believe could be quite common. The scenarios involve an I-O practitioner needing to estimate the criterion-related validity of (a) a single assessment for job X, (b) a battery of assessments for job X with a prespecified weighting scheme, or (c) a battery of assessments for job X with a regression-based weighting scheme. Consider each of these options in turn.[1]

Synthesizing criterion-related validity for a single assessment. In the case of estimating the validity of a single assessment for predicting performance on job X, the practitioner would go to the OSVE and be prompted to enter three pieces of information. The first item required by the OSVE would be ratings of the importance of each element of the GWA taxonomy (or again, some reduced version of them) for the job of interest. The user would have two options here: (a) select an occupation already in the O*NET database (ratings would already be loaded for these occupations in the OSVE) or (b) enter ratings gathered locally. If the user decided to enter information, the Web site would offer stimulus materials (as described earlier) for conducting a structured SME judgment exercise for gathering these ratings to ensure defensibility and consistency of the process (in this case SMEs would likely be supervisors). The second and third items required by the OSVE would be the (a) estimated correlation between the assessment of interest and each element of the GWA taxonomy (or, preferably, a reduced version of them, based on an importance filter) and (b) estimated correlations among individuals' performance on the GWAs in the occupation of interest. Again, the OSVE would allow the practitioner to download stimulus materials for conducting a structured SME judgment exercise for estimating these correlations and to ensure defensibility and consistency of the process (in this case, the SMEs would likely be I-O psychologists). As potential feedback provided to the SMEs, estimated correlations from other users who examined the same occupation could be provided.[2] With the information entered by the user, the OSVE would be able to provide the user with an estimated

1. An anonymous reviewer expressed concern that the OSVE might squelch the use of novel predictor measures. Although we appreciate this reviewer's concern, the options presented here could be successfully executed regardless of the novelty of the predictor measure a practitioner is using as long as the practitioner obtained SMEs who could provide reliable, valid estimates of the relationships between the novel measure and the relevant job components (and perhaps other predictors, should the novel measure be part of a larger assessment battery).
2. Such feedback, if provided, would skew potential efforts to examine relationships among different users' estimates because the estimates provided by later users would partially depend on estimates of earlier users.

https://doi.org/10.1111/j.1754-9434.2010.01254.x Published online by Cambridge University Press

correlation between the assessment and a synthetic performance composite reflecting the combination of GWAs weighted by their importance (or, alternatively, weighted by another set of weights, if desired—unit weights, for example).

Synthesizing criterion-related validity for a battery of assessments with prespecified weights. In the case of estimating the validity for a battery of assessments for performance on job X with a prespecified weighting scheme, the practitioner would go to the OSVE and be prompted to enter five pieces of information: (a) importance ratings for each GWA for the job in question, (b) the estimated correlations among the assessments of interest, (c) the estimated correlations between each assessment and each element of the GWA taxonomy (or some reduced version of them), (d) the estimated correlations among individuals' performance on the GWAs in the occupation of interest, and (e) the weight to be given to each assessment when forming the prediction battery. Regarding the weights given to each assessment, the OSVE would use this information to produce an estimated correlation between the weighted assessment battery and a synthetic performance composite reflecting the combination of GWAs weighted by their importance (per se Nunnally & Bernstein, 1994).

Synthesizing criterion-related validity for a set of assessments with regression-based weights. In the case of estimating the validity for a battery of assessments for predicting performance on job X with a regression-based weighting scheme, the process would be nearly identical to the second option just described. The exception would be that the practitioner would not provide weights; rather, the OSVE would estimate the weights for them. Specifically, with the information entered under Option 2 above, the OSVE could produce an estimated correlation between a regression-weighted assessment battery and a synthetic performance composite reflecting the combination of GWAs weighted by their importance (or, alternatively, weighted by another set of weights, if desired—unit weights, for example). The system would estimate and output the regression weights and estimated validity for the assessment battery based on the underlying matrix containing correlations among the assessments and the synthetic performance composite.

Limitations of the Proposed Approach: Some Issues to Consider

The OSVE would allow practitioners to benefit from the standardized job analysis framework O*NET offers and provide our field with a means of accruing synthetic validity information over time based on a set of standardized stimulus materials and processes. Of course, many issues would need to be addressed to develop the OSVE. Making no effort to offer an exhaustive list, we highlight some of the following issues for consideration.

*Appropriateness of O*NET Data*

Although the Department of Labor's computerized replacement for the *Dictionary of Occupational Titles* holds much promise, the O*NET is not without its critics. One would assume that criticisms levied against the O*NET would commute to any product therefrom derived. Further, despite the expansive nature of the O*NET content model (Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1995), it remains true that others might find additional dimensions that should be added. Johnson et al. provide one example, indicating that the GWA list might need to be expanded to include work behaviors involving citizenship and adaptive performance.

Nevertheless, we believe O*NET largely addresses Johnson et al.'s concern regarding the need to evaluate potential moderator variables such as ''whether the job involves routine or repetitive tasks or whether it operates in a demanding, hazardous environment.'' Although provided just as examples, these variables (and many others) appear in

https://doi.org/10.1111/j.1754-9434.2010.01254.x Published online by Cambridge University Press

O*NET's Work Context scales (#31, ''How often does your current job require that you be exposed to hazardous conditions?''; #42, ''How much time in your current job do you spend making repetitive motions?''; and #51, ''How important to your current job are continuous, repetitious physical activities [like key entry] or mental activities [like checking entries in a ledger]?''). In fact, once the OSVE contained sufficient data, the OSVE database could be used to evaluate whether other elements from the O*NET taxonomy moderate the validity estimates users provide. That is, the OSVE database itself could be a useful data source for enriching what is known about how predictors vary in their relationships to components of performance (as reflected by GWAs) as a function of contextual features of the job.

Pooling SME Judgments

One potentially strong advantage of the OSVE involves the potential for collecting and pooling SME judgments submitted by I-O practitioners who seek validity estimates for various occupations. Given sufficient numbers and representativeness across occupations, these SME validity judgments could be evaluated over time to empirically determine the confidence intervals around mean validity estimates and thereby evaluate one of the hypotheses upon which synthetic validation rests—namely, that predictor/criterion component relationships are essentially invariant across occupations.

Hosting and Development

With any project such as this, ownership is needed. In this case, we agree with Johnson et al. that SIOP would be an ideal host. We see OSVE development as involving both Web development time and time from a working group of I-O psychologist SMEs to develop standard stimulus materials and judgment exercises, as well as the underlying (known) synthetic validation algorithms. In terms of funding its initial development, we do not see the tool as being overly expensive to launch or maintain. Potential sources of initial funding might be sought through a SIOP Small Grant or an appeal to the SIOP Foundation. Given the potential value to our field, we could also see funding for this tool coming from organizations that regularly sponsor SIOP (a fundraising effort in this regard might be undertaken by the SIOP Foundation or Executive Board). We even can envision practitioners and academicians volunteering time (as they do with committees) to help get this effort off the ground and ensure its success.

References

Hough, L. M. (2001). I/Owes its advances to personality. In B.W. Roberts, & R.T. Hogan (Eds.), *The intersection of personality and industrial/organizational psychology* (pp. 19–44). Washington, DC: American Psychological Association.

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

Meyer, R. D., Dalal, R. S., & Bonaccio, S. (2009). A meta-analytic investigation into the moderating effects of situational strength on the conscientiousness–performance relationship. *Journal of Organizational Behavior*, *30*, 1077–1102.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Oswald, F. L., & McCloy, R. A. (2003). Meta-analysis and the art of the average. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 311–338). Mahwah, NJ: Lawrence Erlbaum Associates.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., & Flesihman, E. A. (1995). *Development of prototype Occupational Information Network (O*NET) content model. Volume I: Report [and] Volume II: Appendices*. Washington, DC: American Institutes for Research.

Peterson, N. G., Wise, L. L., Arabian, J., & Hoffman, R. G. (2001). Synthetic validation and validity generalization: When empirical validation is not possible. In J. P. Campbell, & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 411–451). Mahwah, NJ: Erlbaum.

https://doi.org/10.1111/j.1754-9434.2010.01254.x Published online by Cambridge University Press