SYNTHETIC VALIDITY: PAST, PRESENT, AND FUTURE

CHARLES A. SCHERBAUM
Baruch College

Practitioners in the area of personnel selection can be confronted with situations in which the use of traditional validation strategies is not feasible. This situation can result from small numbers of incumbents, rapidly changing jobs, insufficient resources, or the lack of high-quality criterion data. Although traditional strategies are not possible, a set of techniques falling under the rubric of synthetic validity can be used to estimate predictor validity in these situations. Synthetic validity is a logical process of inferring validity on the basis of the relationships between components of a job and tests of the attributes that are needed to perform the job components. Unfortunately, synthetic validity is infrequently used and not well understood. This article provides an overview and conceptualization of synthetic validity, reviews the historical approaches to synthetic validity, describes the recent developments and trends, and suggests future directions for synthetic validity application and research. The relationship of synthetic validity with other validity concepts, the advantages, the limitations, and the legal issues are considered throughout.

A unifying theme in the recent history of personnel psychology has been the movement against the situational specificity of validity (e.g., Schmidt & Hunter, 1977). The advent of meta-analytic validity generalization and subsequent empirical work has convincingly demonstrated that the validity of a predictor or predictor battery can be generalized from one context to another. An equally important development in this movement was synthetic validity. Synthetic validity is an alternative method to establish the generalizability of predictor validity (McCormick, 1959). The advantage of synthetic validity is that it can be used to establish validity evidence in situations where traditional approaches are not feasible (Balma, 1959). However, synthetic validity is infrequently used and is not well understood. In general, it has been a soporific area of personnel psychology. Fortunately, that appears to be changing. There has been a
surge of interest and many exciting developments (e.g., Hoffman, Holden, & Gale, 2000; Jeanneret & Strong, 2003; Johnson, Carter, Davison, & Oliver, 2001) since Mossholder and Arvey’s (1984) review of synthetic validity.

The purpose of this article is to (a) provide an overview and conceptualization of synthetic validity, (b) review the historical approaches to synthetic validity, (c) describe the recent developments and trends, and (d) suggest future directions for synthetic validity application and research. In addition, the relationship of synthetic validity with other validity concepts, the advantages, the limitations, and the legal issues are considered throughout.

Overview and Conceptualization

The term “synthetic validity” was first introduced by Lawshe (1952) at a symposium on industrial psychology for small businesses. Lawshe argued for three concepts of predictive validity: situational validity, generalized validity, and synthetic validity. Situational validity and generalized validity are concerned with the generalizability of a validity coefficient to other situations. Synthetic validity, however, was the inferring of validity in a specific situation. According to Lawshe, validity could be established from a careful assessment of the requirements of a job and judgments about the necessary performance on tests that assess those requirements. Later, at a symposium on synthetic validity at the 1957 meeting of the Midwestern Psychological Association, Balma (1959) refined Lawshe’s definition of synthetic validity to denote, “The inferring of validity in a specific situation from a logical analysis of jobs into their elements, a determination of test validity for these elements, and a combination of elemental validities into a whole” (p. 395). Ghiselli (1959) felt that breaking jobs into their elements would allow us to “synthesize tests into reasonable predictive devices with substantial consistency from job to job” (p. 401).

Ironically, the term “synthetic validity” is a misnomer because it is not actually a type of validity (Guion, 1965, 1998). According to Guion (1998), validity is not synthesized, it is discovered or developed. What this term actually refers to is a logical process of inferring the validity of a test battery from the validities of the tests for the job components that will be included in the battery (Mossholder & Arvey, 1984). Based on these definitions, synthetic validity can be utilized to establish validity evidence using a variety of validation strategies (e.g., content or criterion-related strategies). In fact, a combination of strategies is often used when employing synthetic validity. As noted above, this definition implies that synthetic validity is a form of validity generalization (Jeanneret, 1992; McCormick, 1959). Thus, synthetic validity shares some similarities and assumptions
with other validity generalization approaches (e.g., meta-analytic validity generalization). Comparisons with other validity generalization concepts are discussed in more detail in the subsequent sections.

Synthetic validity was originally developed in response to two needs of the field. The first need was the situational specificity of validity (Ghiselli, 1959). At that time, the consensus in industrial psychology was that validity was situational to the particular organization. Thus, validity needed to be established in every new setting (e.g., Ghiselli & Brown, 1955). Synthetic validity offered a methodology that allowed validities to be used across organizations and jobs. This development represented a valuable and economical addition to the practitioner’s toolbox. However, the importance of synthetic validity in addressing the situational specificity validity has been overshadowed by the developments of meta-analytic validity generalization (Hesketh & Robertson, 1993).

The second need was to develop methods to validate selection procedures when sample sizes were small. This need is an issue of practicality and statistical power. Conducting a traditional validation study is not feasible in small businesses or in jobs with few or no incumbents. If one were conducted, there would be a low probability of detecting a relationship that does exist. By using job components and not jobs, synthetic validity allows one to build up the sample sizes to the levels necessary to conduct validity studies (Brannick & Levine, 2002). For example, a particular job with component “A” may have only five incumbents, but component “A” may be a part of the job for 90 incumbents in an organization. In most cases, the sample sizes will be dramatically larger when using job components. In addition to being a methodological advance, synthetic validity allowed the field to broaden its reach to small businesses. Developing methods to deal with small sample sizes is still a topic of interest to personnel researchers and practitioners (e.g., Hoffman et al., 2000; Johnson, Carter, Davison, et al., 2001; Sackett & Arvey, 1993).

Some researchers have also advocated synthetic validity approaches when criterion data are not available (e.g., Campbell, 1990; McCloy, 1994, 2001). In this case, synthetic validity can be very useful for developing predictor batteries for recently created jobs or for jobs that have not yet been created. The implication is that synthetic validity is an appropriate strategy when the quality of the available criterion data are suspect or in situations where high quality criterion data are very difficult to obtain (e.g., protective services, heavily unionized organizations).

It is important to note that synthetic validity techniques rest on two primary assumptions. The first is that when jobs have a component in common, the human attribute(s) required for performing that component will be the same across jobs. That is, the attribute(s) needed to perform a job component do not vary between jobs. Therefore, a test for a particular
attribute can be used with any job containing the component that requires the particular attribute. The second assumption is that the validity of a predictor for a particular job component is fairly constant across jobs. We would certainly expect variation in a validity coefficient between samples due to unreliability, random error, and other factors. However, this assumption refers to the across-job stability of the population parameter representing the validity of a predictor for a particular job component.

*The Synthetic Validity Process*

The general process of establishing synthetic validity consists of three steps (Guion, 1965; Hoffman & McPhail, 1998; Mossholder & Arvey, 1984). In the first step, a structured job analysis is conducted to determine the important components of a particular job or job family. The job analysis can be either task or worker orientated, but the metric of the analysis should allow for across-job comparisons (McCormick, 1959). In addition, the job components should be described at a level of detail that will facilitate the identification of predictors to assess the job components. Given that the job components are the basis of establishing evidence of the validity of a predictor, it is essential that the results of the job analysis are accurate and reliable. Errors in the job analysis can easily lead to errors in the subsequent selection of predictor tests (Mossholder & Arvey, 1984).

The second step involves selecting predictors and then estimating the relationship between job components and predictors of those components. Predictors are typically selected using subject matter experts' (SME) judgments or previous research that demonstrates a predictor test is related to the job component. When SMEs are used, they can be asked to rate, for example, the criticality of a predictor test of a particular attribute for performing a particular job component. The predictors that are rated to be the most critical are retained in the subsequent steps. As described here, the test selection process assumes that predictor tests are already available that can be used to predict performance on the job components (Mossholder & Arvey, 1984). However, new predictors can also be developed. In these cases, empirical work would be necessary to demonstrate the psychometric properties of new predictors.

The relationships between the predictors and job components can be established using a variety of validation strategies (Guion, 1976). As noted above, multiple strategies are often applied. In fact, synthetic validity is really a specific methodology for employing these generalized strategies for establishing validity. For example, one can use a traditional criterion-related strategy to derive estimates of the relationship between the predictors and the job components (Arvey, 1979; Guion, 1965). To use this strategy, a concurrent validation study is typically used. Objective
or subjective performance criteria for the job components are correlated with the scores from the predictors that were selected to assess the job components. This process is carried out for each predictor–job component pair. Alternatively, the results of previous validation studies could be used (Schneider, 1976).

Most often, the strategy employed relies on SME judgments to establish the relationship between the predictor(s) and job component(s). When using SME judgments, the relationship between the predictors and job components is estimated indirectly. For example, SMEs can be asked to rate the importance of a predictor for a job component. The importance ratings are averaged across several raters and then divided by the number of points on the rating scale to create a proportion. This proportion can be used as an estimate of the relationship between the predictor and job component (Hamilton & Dickinson, 1987). Alternatively, policy capturing could be used to derive an estimate of the relationship between the predictor(s) and job component(s). In some cases, particular SMEs (e.g., I-O psychologists, test experts) are asked to directly estimate the empirical relationship between each job component and predictor.

This second step highlights several important aspects of synthetic validity. First, it highlights some of the advantages of synthetic validity. Synthetic validity requires that predictors be linked to specific components of job performance instead of global assessments of job performance (Dunnette, 1966; Ghiselli, 1959) and it provides flexibility in establishing these linkages. Second, it highlights the two assumptions underlying the synthetic validity noted above: when jobs have a component in common, the human attribute(s) required for performing that component will be the same across jobs; the validity of a predictor for a particular job component is fairly constant across jobs. Third, it highlights the role of SME judgments in synthetic validity. All synthetic validity approaches use SME judgments to some degree, but the methods clearly differ on how and at what point SME judgments are used. Some methods only use SME judgments in collecting job analysis or performance data but other methods use SME judgments as the basis of the validity estimates. The role of SME judgments is noted in the discussion of each of the major synthetic validity methods.

The third step involves forming a test battery for a job using the validity information from the previous step. This step entails estimating the expected validity of the test battery. The estimated validity is a weighted composite of the validities for each predictor in the battery. That is to say, the estimated validity is computed using some form of a linear equation. The data entered into these equations depend on the particular synthetic validity approach but typically include the estimates of the predictor-job component relationships, ratings about the job components, or scores from
the predictors. There are a variety of empirical (e.g., regression) or rational (e.g., SME importance ratings) weighting schemes that can be used (see Peterson, Wise, Arabian, & Hoffman, 2001 for a discussion of different weighting schemes) and the scheme that is chosen will often depend on the particular synthetic validity method that is used. In addition to estimating the validity, some of the synthetic validity approaches (e.g., job components validity) use these equations to estimate the mean scores on the predictor tests.

This process assumes that overall job performance is a function of the performance on job components (Wise, Campbell, & Arabian, 1988). Although many would agree with this assumption, its veracity should be considered in each application of synthetic validity. If a job cannot be broken into its components, then it is not appropriate to use synthetic validity to develop validity evidence for that job. Given that jobs can change rapidly, the ability to create test batteries based on job components is a major advantage of the synthetic validity approach.

Figure 1 provides an illustration of this basic process as described by McCormick (1959). On the left-hand side of the figure, there are four job components (1, 2, 3, and 4) and four jobs (A, B, C, and D). Stars in the cells indicate that the component is part of the job. There are also four predictor tests (W, X, Y, Z) for the job components that have some form of validity evidence. For jobs A and D, a test battery consisting of tests W and Z could be formed because these jobs contain components 1 and 4. The validity of this test battery could be synthesized from the validities of test W and Z. Two new jobs (E & F) are created. A job analysis reveals that job E, for example, will contain components 2 and 4. A test battery could be formed using tests X and Z and the previously established validities of these tests could be used to estimate the validity of this test battery.

**Historic Approaches to Synthetic Validity**

Synthetic validity is best described as a family of methods for inferring predictor validity. In fact, numerous approaches to and perspectives on synthetic validity have been reported in the literature (Guion, 1976, 1998;
Mossholder & Arvey, 1984). This section briefly reviews the techniques developed prior to 1990. More recent developments are discussed in the subsequent sections. Although each technique contains the three general steps outlined above, they are conceptually different. The conceptual logic and procedures of each technique are discussed.

**J-Coefficient**

The first approach is the J-coefficient (Primoff, 1955, 1957, 1959). The J-coefficient is a worker-orientated approach to analyzing jobs and predictors in terms of the employee characteristics that lead to superior performance using the job elements method of job analysis. This analysis leads to the identification of the aspects common to multiple jobs (i.e., job components) and the predictors, which in turn, are used to estimate the test battery-job performance relationship. Specifically, this information is used to create the J-coefficient, which mathematically indexes the test battery and job performance relationship. Several authors have argued that this approach is similar to a content validation strategy (e.g., Guion, 1976; Mossholder & Arvey, 1984). However, it also draws on a criterion-related validation strategy.

There are many formulas that can be used to compute the J-coefficient (see Hamilton & Dickinson, 1987 for a comparison of the formulas). A common J-coefficient formulation presented by Hamilton (1981) is as follows:

\[
J_{xy} = \frac{R_{xe} R_{ee}^{-1} R_{ye}}{\sqrt{R'_{xe} R_{ee}^{-1} R_{xe}} \sqrt{R'_{ye} R_{ee}^{-1} R_{ye}}}
\]

where \( R_{xe} \) is a vector of the relationships between the predictors and job components, \( R_{ee} \) is a matrix of the relationships between the job components (in some formulations this term is not required, see Hamilton & Dickinson, 1987), \( R_{ee}^{-1} \) is the inverse of the \( R_{ee} \) matrix, and \( R_{ye} \) is a vector of the relationships between the job components and job performance. The formulas for the J-coefficient are based on the assumptions of linearity and will typically underestimate the “true” test validity. However, correction formulas are available (Hamilton, 1981).

There are multiple ways to estimate \( R_{xe} \) and \( R_{ye} \). First, there can be estimates using SME judgments. For example, \( R_{xe} \) can be estimated using SMEs’ or job analysts’ ratings of the relevance of each test item for each job component. The proportion of items judged to be relevant can be used as the estimate of \( R_{xe} \) (Hamilton, 1981). Alternatively, SMEs can be asked to rate the importance or relevance of a predictor for a job component. The ratings of importance or relevance can be averaged across raters and then
divided by the number of points on the rating scale to create a proportion. This proportion is used as the estimate of $R_{se}$. Alternatively, the $R_{se}$ vector could be derived from test experts or I-O psychologists’ estimates of the validity coefficients. To estimate $R_{se}$, the ratings of SMEs or job analysts of the importance or contribution of each job component to overall job performance can be used (following the procedures described above for $R_{se}$). Policy-capturing procedures can also be used to estimate $R_{se}$ with SMEs’ ratings (Hamilton & Dickinson, 1987).

Second, provided large samples and databases of test scores and performance ratings (both component and overall), $R_{se}$ and $R_{se}$ can be estimated empirically using the correlation between the predictors and the job components in the case of $R_{se}$, and the correlation between the measures of component performance and overall performance in the case of $R_{se}$ (Primoff, 1959). In practice, however, SME judgments are typically used. The question then becomes which rating source will provide the most accurate estimates. Hamilton and Dickinson (1987) compared J-coefficients derived from several sources including the ratings from supervisors, incumbents, coworkers, and test experts with criterion-related validities. They found that no source was overwhelmingly superior. They recommend that the choice of source could be driven by practical considerations. Evidence of the J-coefficient technique in practice is limited (Mossholder & Arvey, 1984). Primoff (1959) reported test validity estimates from the J-coefficient method that ranged from $.23$ to $.62$. Hamilton and Dickinson (1987) and Trattner (1982) have both presented evidence that J-coefficient estimates are comparable to traditional validity coefficients.

The advantage of this technique is that the process can potentially involve many stakeholders (e.g., incumbents, supervisors, union officials), which may possibly lead to increased acceptance of the eventual selection system. The resources (e.g., cost, time, expertise, and personnel) required to use this technique are the major drawback of the J-coefficient method. As a result, this approach has been used primarily with large employers and the government. However, it is still possible to use this method in some small and medium-sized organizations.

Another concern with this method is that it uses statistical prediction, which is based, in many cases, solely on SME judgments. Although the use of expert judgment is not inherently problematic (e.g., Dawes, 1979; Sawyer, 1966), it does raise some issues. First, it may be difficult to get the data needed for the validity estimates from some types of SMEs. For example, supervisors may not be familiar with some aspects of the job, and therefore, cannot provide ratings for those job components (Mossholder & Arvey, 1984). Moreover, some types of SMEs may not be capable of making these types of estimates and judgments. Second, this method is based on improper linear models (Dawes, 1979) because it uses
nonstatistical techniques to determine the weights (i.e., \( R_{xe} \) and \( R_{ye} \)) in the J-coefficient equations. For example, in some cases this method uses test experts and I-O psychologists to directly estimate validity (i.e., \( R_{xe} \)). There is evidence supporting the capability of these individuals to make these judgments (e.g., Schmidt, Hunter, Croll, & McKenzie, 1983) and it is necessary in some situations (Dawes, 1979; Guion, 1998). Nevertheless, it is a form of expert judgment and cautions about using this type of data clearly apply (e.g., Mehl, 1954). More recently, Primoff has developed techniques that minimize this concern (e.g., Primoff, 1980). Although the J-coefficient was the approach of choice for many personnel specialists early on, interest now has waned (Hamilton, 1981).

Job Component Validity

The second approach is job component validity (McCormick, 1959). Job component validity (JCV) has been the dominant approach to synthetic validity. McCormick (1959) outlined this technique over 40 years ago, but the theoretical development was ahead of the necessary technological developments (i.e., the Position Analysis Questionnaire [PAQ]). Conceptually, this technique is similar to a criterion-related validation strategy (Guion, 1976). However, the validity coefficients are predicted instead of computed as is done in a traditional criterion-related validation strategy (Mossholder & Arvey, 1984). Therefore, the tests and performance are linked indirectly (Trattner, 1982). The logic behind the JCV approach is that if one could identify the human requirements for any given component, it would be possible to determine the total requirements for any job from knowing the components of that job (McCormick, 1976; McCormick & Mechan, 1970).

At its core, the JCV method is a set of equations for predicting mean test scores and validity coefficients from job dimension scores. These equations were empirically derived using test scores from the General Aptitude Test Battery (GATB) and validity coefficients from validation studies by the U.S. Employment Service (Jeanneret, 1992; McCormick, 1976; McCormick, Jeanneret, & Mechan, 1972). In these equations, mean tests scores and validity coefficients are the dependent variables and job dimension scores, such as those from the PAQ, are the independent variables. Although the original development of JCV relied on the GATB and U.S. Employment Service data, research has extended this technique to predicting mean scores and validity coefficients for commercially available tests (e.g., McCormick, DeNisi, & Shaw, 1979).

When the PAQ dimension scores are used to predict mean test scores, it assumes that individuals gravitate to jobs in which the requirements match the individuals’ ability (e.g., high-ability individuals gravitate toward jobs
with high-ability requirements and low-ability individuals gravitate toward jobs with low-ability requirements; McCormick et al., 1979). Thus, the rank ordering of jobs based on the test scores of the incumbents in those jobs roughly reflects the ability requirements for those jobs. Some authors have referred to this assumption as the “gravitational hypothesis” (Sparrow, 1989; Wilk, Desmerais, & Sackett, 1995). Although there is empirical evidence supporting the notion of individuals gravitating toward jobs with ability requirements that match their level of ability (e.g., McCormick et al., 1979; Sparrow, 1989; Wilk et al., 1995), it has been a source of criticism.

Operationally, the JCV technique consists of four steps (Mossholder & Arvey, 1984). First, a job analysis is conducted using the PAQ. Second, the PAQ ratings from the SMEs are used to determine the components that are part of a given job. Third, regression analyses (based on equations from previous JCV studies as noted above) are used to weight the component scores and predict the mean scores of the incumbents in that job for the GATB tests or predict the validity coefficients. Fourth, the results from the regression analyses are used to form a battery of tests measuring the attributes necessary to perform the job components in a particular job. More technical details of this strategy can be found in McCormick, Cunningham, and Thornton (1967), McCormick et al. (1972), and the PAQ user’s manual (McCormick, Mecham, & Jeanneret, 1989).

McCormick and colleagues (McCormick et al., 1967, 1972, 1979; Sparrow, 1989; Sparrow, Patrick, Spurgeon, & Barwell, 1982) have provided evidence of the usefulness of this technique. The multiple correlations between the job components and the validity coefficients or the mean test scores have ranged from near .00 to near .85. The cross-validated multiple correlations have ranged between .59 and .80 when predicting mean test scores and between .40 and .55 when predicting validity coefficients (see McCormick et al., 1972). The size of the coefficients, in general, has been criticized as being smaller than would be expected (Sackett, 1991). The prediction for mean GATB test scores is better than the prediction of the validity coefficients. As argued by Sackett (1991), this result is not surprising because means are more stable point estimates than correlations. Means are based only on a univariate distribution and are not subject to distortion by artifacts such as sampling error or range restriction. On the other hand, the size and variability of the validity coefficients may be a result of the small sample sizes in the U.S. Employment Service’s validation studies and corresponding methodological artifacts (e.g., sampling error; McCormick et al., 1989).

The JCV technique has clear advantages in terms of ease and cost of use. This method of synthetic validity can be easily used in large and small organizations alike. As will be described below, a second advantage
is that several efforts have been made to incorporate the JCV approach into models of construct validity. In addition, the JCV does not use SME judgments to estimate the relationships between the predictors and job components or the weights in the predictive equations. In both cases, the relationships are estimated empirically. In light of its advantages, the use of the JCV approach has steadily grown (Hoffman & McPhail, 1998; Jeanneret, 1992; Sparrow, 1989). The drawback of this technique is that it provides little detailed information about the component–performance or predictor–component relationships. With detailed information on the job components, it is relatively easy to infer the attributes needed to perform the job components. The reverse is not true. It is by no means a simple task to infer the job components from worker-orientated job analysis data. As a result, cross-organization synthetic validity efforts will be difficult if some of the organizations used “task-oriented” job analysis methods and others used “worker-orientated” job analysis methods.

Guion’s (1965) Approach

The third approach is a method presented in a study by Guion (1965). This approach differs from the others in that it uses a traditional validation strategy to establish predictor validity (Guion, 1976). Validity is estimated based on the empirical relationships between the measures of performance on the job components and the predictors for those components. As with the JCV, the estimation of validity of the test battery is not based primarily on SME judgments. However, this method estimates validity more directly than the JCV or J-coefficient approaches in that it directly links the predictors and criteria.

To demonstrate this approach, Guion first conducted a job analysis of the various jobs in a small organization. He determined that the jobs contained combinations of seven different components. Component and overall performance ratings were obtained for each employee. The employees were then given a battery of selection tests. The most important predictors for each job component were identified by examining the differences in test scores for each predictor among the employees that were rated as successful on the particular job component. The multiple correlations between the two most important predictors and the job component performance ratings were then computed. Based on the empirical relationships between the test scores and performance ratings for the components, expectancy tables were developed for use in future employee selection. The expectancy tables indicated the likelihood of an employee receiving a rating of superior on a job component for different ranges of scores on two predictor tests.
The major strength of this approach is that it directly estimates the validity of a test battery. That is, this approach produces empirical estimates of the relationship between the predictors and job components. Moreover, this approach is very similar to traditional validation strategies. Thus, it may be easier to use than other approaches. Nevertheless, subsequent research using this exact strategy has not been reported, but the rationale of this approach is apparent in several recent synthetic validity applications.

Hollenbeck and Whitener’s (1988) Approach

The fourth is an approach developed by Hollenbeck and Whitener (1988). They argue that their approach is a combination of the J-coefficient approach, the JCV approach, and Guion’s (1965) approach. The major difference is the order of aggregation of the validity estimates. In the J-coefficient, JCV, and Guion (1965) approaches, the test–job component relationships are assessed first and then aggregated. In Hollenbeck and Whitener (1988), the performance ratings on the job components and scores on the predictor tests for the components are aggregated and then the relationship is assessed. Thus, for each employee an aggregated test score and an aggregated component performance score is computed. The relationship between predictors and component performance is then assessed across all employees. As a result, the sample size in Hollenbeck and Whitener’s (1988) approach is the number of employees in an organization instead of the number of employees sharing a job component. Thus, this approach allows for the inclusion of jobs with smaller numbers of incumbents than is possible with the other synthetic validity approaches.

To demonstrate their method, Hollenbeck and Whitener (1988) conducted a job analysis using the PAQ. From the PAQ ratings by SMEs, the job components were identified. A test battery measuring the attributes required for these components was administered and performance ratings on these components were collected. The test scores and performance ratings were then aggregated for each employee. Finally, the relationship between the aggregated test scores and the aggregated component performance ratings across all employees was assessed. In fact, they found there was a relationship between the aggregated test and performance scores, albeit one that was moderate in magnitude. As is true of Guion’s (1965) method, the strength of this approach is that it directly estimates the validity of a test battery. Despite the methodological and statistical sophistication of their method, it has had little impact on the field to date. This situation is unfortunate given that this approach is very useful for small organizations. Moreover, several authors have advocated the use of this approach (e.g., Schneider, Hough, & Dunnette, 1996).
Other Approaches

There are two other approaches to synthetic validity that are worth noting. Each of these techniques is much less developed than the J-coefficient, the JCV, the Guion, or Hollenbeck and Whitener approach. The first is Lawshe and Steinberg's (1955) clerical checklist. Interestingly, this method makes the same gravitational assumption that is used when predicting mean test scores in the JCV approach. This method consists of supervisors rating how critical the abilities on the Purdue Clerical Adaptability Test are for the performance of each clerical operation on the Job Description Check List of Clerical Operations developed by Culbertson (1947). This checklist contains 139 different clerical operations that were identified by Culbertson. A test battery was formed from the abilities that were rated to be the most critical by the supervisors. This battery was then administered to incumbents and expectancy charts were developed that reflected the probability of incumbents exceeding the median score on each of the tests given the number of critical operations that are part of a particular job. These expectancy charts could then be used on other jobs that contain the same critical operations. Although Lawshe and Steinberg (1955) found partial support for this approach to synthetic validity, no subsequent work using this method has been reported.

The second method is the functional job analysis (FJA; Fine, 1963). In fact, Sidney Fine's doctoral dissertation was an attempt to develop a synthetic validity approach based on FJA (Fine, 1962). This method is similar to the JCV method. The basic strategy is to establish the dimensions of a job and their importance within the hierarchical people, data, and things classification. Then, the human requirements for the dimensions are established using the ratings on the FJA dimensions to predict mean test scores. Finally, a test battery can then be synthesized using the dimensions of the job. Initial work was supportive of this approach, but no efforts to further develop this method have been presented since Olsen, Fine, Meyers, and Jennings (1981). The FJA approach possesses many of the same advantages as the JCV method and it can be applied in large and small organizations. However, this approach is far less developed than the other approaches and much less guidance is available for researchers and practitioners who wish to use it.

Most of the approaches described above have primarily involved building cognitive test batteries. However, the synthetic validity approach has been utilized in a few instances with work samples and psychomotor ability tests (Griffin, 1959). For example, Drewes (1961) presented a synthetic validity approach to creating psychomotor test batteries from an analysis of the specific motion patterns of a particular job. This approach involved matching the specific motions to tests of those specific motions. He found
that his approach resulted in validities that were superior to general psychomotor motion tests. Kintop and Mussio (1974) have utilized synthetic validity approaches to develop work sample batteries to assess the critical job components in clerical jobs. Although conceptually appealing, developing work samples for each job component may not be practical given the cost and time that are required to undertake such an effort (Mosholder & Arvey, 1984).

Current Developments in Synthetic Validity

In their review, Mosholder and Arvey (1984) concluded that research, at that point, had demonstrated the potential usefulness of synthetic validity, but the concept had yet to have a substantial impact on the field of personnel selection. They argued that the research left many questions and had not captured the interest of personnel specialists. Although still largely true, the situation is changing. There has been a renewed interest in synthetic validity and several new developments have occurred. Below, a number of these developments are discussed.

Many of the recent developments in synthetic validity methods have occurred in the military. The most notable is the Army’s Synthetic Validity (SYNVAL) project (Peterson et al., 2001). This project represents one of the largest synthetic validity studies ever conducted. In this project, multiple methods of developing (i.e., weighting) synthetic validity equations were used and the synthetic validity estimates were compared to validity estimates derived from empirical studies (i.e., Project A). The synthetic validity process utilized in this project closely resembled the four steps outlined at the beginning of this article and in Guion’s (1965) approach in many respects.

Drawing on preexisting performance taxonomies (e.g., Fleishman & Quaintance, 1984) in addition to job analysis data, the job components of several military occupational specialties (MOS) were determined. Predictors were drawn from the test batteries that were developed as part of Project A. The scores on these predictors were used as the independent variable in the synthetic validity equations. The magnitude of the relationship between the job components and predictors were determined using the judgments of I-O psychologists. These estimates were used as the basis for the different weighting schemes in the equations to predict core technical proficiency and overall performance in each of the MOS that were studied.

It was found that all of the methods for developing synthetic validity equations worked well and that there was little difference between the validity coefficients derived from the synthetic validity study and validity coefficients derived from the Project A data for predicting core technical
proficiency and overall performance. Peterson et al. (2001) also examined the discriminant validity of the synthetic validity equations. That is, how well does the synthetic validity equation developed for one MOS predict the core technical proficiency and overall performance for different MOS? In general, the synthetic validity equations demonstrated very little discriminant validity. The mean differences in validity when the equations were used for the MOS it was developed for and when they were used for different MOS were very small (i.e., \( M = -0.01 \) to \( M = 0.06 \)). However, this finding could be a result of the large number of cognitively oriented predictors in the test batteries (e.g., verbal ability, numerical ability). As Schmidt and Hunter (1998) have argued, cognitive ability is a good predictor of job performance for a wide variety of jobs (i.e., it lacks discriminant validity across most jobs).

An interesting aspect of this project was that it provided further empirical support for the usefulness and appropriateness of using expert judgments in validation research. I-O psychologists were used to estimate the relationships between the predictors and job components. This finding is important given that some synthetic validity approaches (e.g., J-coefficient) rely on SME judgments to develop the validity equations. Nevertheless, this method does rely on expert judgments to determine the weights in the regression equations (i.e., improper linear models; Dawes, 1979). Again, the use of expert judgment is not inherently problematic, but concerns about this type of data may apply (e.g., Dawes, 1979; Meehl, 1954).

Building on the SYNVAL project, McCloy (1994, 2001) proposed a new analytic approach for establishing synthetic validity equations. Specifically, he has proposed the use of multilevel modeling as a means to link individual and job-level characteristics to make predictions about individual performance (see Hofmann, 1997 for an introduction to multilevel models). For example, these equations can be used to model the effects of job components (Level 2) on the relationship between predictor and criterion measures (Level 1). The general multilevel equations can be simplified to job-specific equations based on SME ratings (e.g., level of importance) of the job components for that job. Thus, job-specific predictions can be made for individuals in any job that contains the set of components in the equations.

Using data from Project A and the Dictionary of Occupational Titles, McCloy (2001) demonstrated that the multilevel modeling approach produces results that are comparable to the results from job-specific ordinary least squares regression analyses and the predictive equations developed from the SYNVAL project. Specifically, McCloy used several different test scores (e.g., AFQT) and experience measures (e.g., time in service) at Level I and the level of several job characteristics (e.g., working with
things) at Level 2 with the scores from a hands-on performance test as the criterion at Level 1. Although McCloy developed his equations using Project A data, he notes that multilevel equations can be developed from other types of job analysis data (e.g., O*NET) and the ratings of SMEs. The combination of a synthetic validity approach and multilevel modeling offers a great deal of flexibility in empirically generating general predictive equations that can be tailored to any specific job containing the job components in the multilevel model. Despite the appeal of this approach, it requires sample sizes that are much larger than what may be found in the typical small organization. Thus, it may be better suited for large organizations, the military, or industry consortia. Nonetheless, this combination represents a substantial advance in synthetic validity methodology.

Hoffman and colleagues (Hoffman & McPhail, 1998; Hoffman & Morris, 2003; Hoffman et al., 2000; Morris, Hoffman, & Schultz, 2003) have recently reported several studies that combine synthetic validity and other validity generalization approaches. Hoffman and McPhail (1998) address the question of the similarity of the validity estimates from synthetic validity and meta-analytic validity generalization (VG). Campbell (1990) speculated that they should be similar, but at the time there were no data available to test this speculation. Employing the JCV approach to synthetic validity, Hoffman and McPhail (1998) used archival data to test the comparability of these techniques.

Drawing on the VG estimates for clerical jobs published by Pearlman, Schmidt, and Hunter (1980) and a PAQ database from a large utility company, Hoffman and McPhail (1998) found substantial correspondence between the validity estimates from the JCV and VG approaches. In fact, they found a correlation of .97 between the uncorrected mean VG and mean JCV estimates for clerical occupational groups using cognitive ability tests. The JCV approach tended to produce coefficients that were the same or slightly more conservative than the uncorrected VG estimates. They argue that these results provide compelling initial evidence of the similarity of synthetic validity and VG validity estimates. Further, because JCV is more firmly rooted in job analysis and is applicable to any job, it may be a better choice than VG in some situations.

Morris et al. (2003) extended the research of Hoffman and McPhail (1998) by including additional jobs (i.e., nonclerical) in their comparison of validity estimates derived from a JCV study and those obtained from

---

\(^1\)This study and the following studies raise an important caveat about the comparison of VG and synthetic validity estimates. Typical VG estimates (i.e., fully corrected) will always be larger than synthetic validity estimates given the corrections. Nevertheless, the difference between the VG estimates (corrected and uncorrected) and synthetic validity estimates tends to be small.
a meta-analysis of validity coefficients collected by commercial test publishers. As was found by Hoffman and McPhail (1998), the JCV estimates were highly similar to the uncorrected meta-analytic validity estimates. The majority of the JCV estimates fell within the 95% confidence interval around the meta-analytic estimate. Again, the synthetic validity estimates tended to be more conservative than the meta-analytic estimate. In sum, both of these studies demonstrate that synthetic validity compares favorably to the current gold standard of meta-analytic validity generalization.

In addition, Morris et al. (2002) proposed a new method for combining individual JCV validity coefficients to estimate the validity coefficient of a battery of predictors. This approach utilizes an intercorrelation matrix of the subtests from either the GATB or commercially available tests. Using these intercorrelation matrices, a multiple correlation between the predictors and a criterion (e.g., overall job performance) is computed. Hoffman and Morris (2003) used this procedure to compare JCV estimates for a test battery with empirical validity estimates for the same test battery that were corrected for range restriction and unreliability. They found that in all cases the JCV estimates were comparable to the empirical estimates, but the JCV provided a conservative estimate of test battery validity. However, the differences between the estimates were small, particularly in comparisons with the uncorrected validity estimates.

In another study, Hoffman et al. (2000) combined JCV, cluster analysis, and previous VG results to establish a test battery and cut scores for jobs in a large utilities company for which a selection battery did not exist. In this case, a synthetic validity approach was used because there were too many jobs titles for which validity evidence was needed. Using an existing database of PAQ ratings, job families were formed from a cluster analysis and the JCV validity estimates for the job family were compared to validity estimates from previous validation studies. From the comparison of validity estimates, they found that the predicted validities from the JCV method were very similar to the uncorrected observed validities. Again, the JCV approach tended to produce more conservative validity estimates. Taken together, this series of studies supports the similarity of estimates derived from synthetic validity, VG, and traditional validity studies. Moreover, they demonstrate the usefulness of synthetic validity when the number of job titles requiring validation studies exceeds an organization’s capability (e.g., cost, time, personnel) to conduct such studies.

Johnson and colleagues have also presented several new developments and applications of synthetic validity. Johnson, Carter, Davison, et al. (2001) extended synthetic validity techniques to the testing of differential prediction hypotheses. Specifically, they developed a method called synthetic differential prediction analysis that uses the logic of synthetic validity to examine if a predictor test results in differential predictions for
individuals in different subgroups (e.g., race, sex). Typically, this hypothesis is difficult to test because of small sample sizes for many subgroups, and thus, the power to detect differences is low. Using job components, however, one can build up the sample sizes for various subgroups.

As an illustration of the power of this strategy, they presented an example where the sample size for a white-black within-job family comparison was 0 at the job family level, but the sample size was 703 at the job component level. For other comparisons, the changes in sample size exceeded 1000. Ironically, a potential limitation of this method is that the statistical tests could become too powerful. Therefore, these authors advocate the use of effect sizes with this technique. They also caution that an effect size considered to be trivial (e.g., $\Delta R^2 < .01$) by the researcher may not be seen as such in court. This technique potentially offers a means to test differential prediction hypotheses in small organizations or for jobs with few incumbents.

Johnson, Carter, and Tippins (2001) extended the synthetic validity approach to situations where validity coefficients are needed at the job family level. Their approach is similar to the approach presented by Hoffman et al. (2000). The major difference is that the extension presented by Johnson, Carter, et al. (2001) is rooted in the synthetic validity approach developed by Guion (1965), instead of the JCV approach.

Johnson, Carter, et al.'s (2001) approach first involves conducting a job analysis to identify the job components, job families, and predictor constructs. Using this information, measures of the predictor constructs and performance ratings for the job components are developed. Data on the predictors and criteria are collected, and SME judgments are then used to identify the most relevant predictors for the job components. Based on the components common to a job family, a predictor battery is then formed. The validity coefficient between the predictor battery and a composite of the job component performance ratings is computed using Nunnally's (1978) formula for computing the correlation between two composites.

Johnson, Carter, et al. (2001) demonstrated this approach in a large telecommunications company that wished to develop a selection system for screening a large number of entry-level jobs. As was the case in Hoffman et al. (2000), a synthetic validity approach was used because validity evidence was needed for an exceptionally large number of job titles. In addition to conducting the synthetic validity study, data were collected within job families that could be used to compute empirical validity coefficients. As a result, they were able to compare the deviations between the synthetic and traditional validity coefficients. In general, the empirical coefficients were larger than the synthetic validity coefficients. However, when using a 90% confidence interval, there were no differences between the empirical and synthetic coefficients. Johnson, Carter, et al. (2001)
concluded that the cost and time savings in conjunction with the similarity to traditional validity estimates make this synthetic validity approach an attractive option. Because this approach is at the job family level, it is particularly useful when there are too many job titles that need validity evidence. They further argue that in environments where new jobs are rapidly added and jobs are constantly changing, their approach offers a flexible alternative to traditional validity studies.

Most recently, there have been several efforts to link synthetic validity and the O*NET database and the Dictionary of Occupational Titles (DOT). Although the DOT is being replaced by O*NET, several studies have drawn on the DOT database in synthetic validity efforts. McCloy (2001) utilized ratings from the DOT in developing his multilevel modeling approach to synthetic validity. Recently, Wagner and Harvey (2005) have used a JCV approach to link the work dimension scores from the Common Metric Questionnaire (Harvey, 1991a) to worker-requirement ratings for occupations listed in the DOT.

Jeanneret and Strong (2003) have presented a study linking the JCV approach to the O*NET database. Up to this point, the JCV approach utilized the job component ratings from the PAQ. Using both rational and cross-validated empirical linking strategies, Jeanneret and Strong demonstrated that the generalized work activities ratings from O*NET can predict mean test scores on the GATB and the Wonderlic cognitive ability test. The magnitude of the observed correlations was similar to those seen in previous JCV research that uses the PAQ as the job analysis instrument. In a similar study, Johnson, Carter, and Dorsey (2003) also used the JCV approach to predict mean GATB scores from the O*NET database. However, in addition to the generalized work activities, they included data from the knowledge, skills, and abilities domains in the O*NET database. They also found that the O*NET data were predictive of the mean GATB scores and the magnitude of the relationships were similar to those seen in previous research with the PAQ.

In another study, D'Egidio (2001) used the JCV approach to predict mean test scores on commercially available tests (e.g., Workplace Literacy Tests, Hogan Personality Inventory) from the skills and generalized work activities data in O*NET. She found that the relationships between the O*NET data and the mean test scores on the commercial predictors were similar to those seen in JCV research that uses the PAQ and GATB test scores. Wagner and Harvey (2004) used a JCV approach to predict mean scores for the 52 ability traits from Fleishman's (1975) ability requirement scales that are included on the O*NET ability survey (Fleishman, Costanza, & Marshall-Mies, 1999) from the generalized work activities ratings in the O*NET database and SME ratings on Common Metric Questionnaire (Harvey, 1991a). They found that the relationships between mean ability
scores and both the O*NET and Common Metric Questionnaire data were similar to those seen in previous JCV research using the PAQ and the GATB. Although more research is clearly needed to further explore the linkage between O*NET and synthetic validity, the findings of these initial studies suggest that synthetic validity approaches have the potential to be applied to almost any job in the U.S. economy. Moreover, this development further increases the accessibility of synthetic validity approaches to small organizations.

Lastly, there have been some developments in techniques for linking ability requirements to job components. Barrett and Driskill (1990) and Earles, Driskill, and Dittmar (1996) have presented a system to link ability requirements to semantic task categories based on SME ratings. Specifically, task categories are created using the action verbs in the task statements. Thus, each category has one action verb (e.g., analyze) and multiple objects (e.g., records, data, communications). The importance of a standardized set of abilities (e.g., Fleishman & Mumford, 1988) is then rated by SMEs for the action verb—object pairs in each semantic task category. Based on the ratings of the SMEs, predictor tests for the most important abilities in each semantic task category can be identified. Test batteries can then be formed based on the semantic task categories that are part of a particular job. As noted by Earles et al. (1996), empirical validation of the linkages is necessary to evaluate the usefulness of the linkages. Given the large number of task statements that can be generated in a task-orientated job analysis, this approach provides a useful approach to identifying job components and linking job components to abilities. However, because an empirical evaluation of this technique has not been presented, the ultimate usefulness of this technique for synthetic validity approach cannot yet be determined.

Advantage of Synthetic Validity

As noted throughout this article, synthetic validity regardless of the specific approach offers several advantages. The initial body of research has clearly demonstrated that synthetic validity is useful in numerous contexts and that the estimates derived from synthetic validity approaches are conservative but comparable to empirically derived estimates. With the incorporation of O*NET into the synthetic validity framework (e.g., D'Egidio 2001; Jeanneret & Strong, 2003; Johnson et al., 2003), conducting a synthetic validity study is becoming even less onerous and costly to organizations. Because of the importance of these advantages, two of the overarching advantages are briefly reiterated.

In terms of practicality, all of these approaches offer a strategy for establishing estimates of the validity of predictors when traditional test
validation strategies are not feasible. Synthetic validity approaches can be used in small organizations, with jobs containing few incumbents, with rapidly changing jobs, with jobs that were recently created, or with jobs that do not yet exist. Moreover, because most jobs can be broken down into the major components, synthetic validity approaches can be applied to develop a test battery for almost any job (Dunnette, 1966; Jeanneret & Strong, 2003). As demonstrated in the work of Johnson, Carter, et al. (2001) and Hoffman et al. (2000), synthetic validity is equally useful when there are too many job titles for which validity evidence is needed. Thus, it offers a flexible strategy for predictor validation and a valuable addition to the practitioners’ toolbox.

Many authors have bemoaned the dominance of global and univariate conceptualizations of performance criteria (e.g., Campbell, 1990; Murphy & Shiarella, 1997; Sells, 1966). Murphy and Shiarella (1997) argue that the univariate conceptualizations are not a realistic reflection of organizational selection practices and call for the adoption of multivariate approaches. The use of synthetic validity requires a componential and multivariate view of performance. In essence, synthetic validity compels researchers and practitioners to adopt a construct-oriented approach to criteria development and predictor selection (Schneider et al., 1996). In all synthetic validity approaches, predictors are selected to specifically match particular components of a job. Selecting predictors in this manner is in contrast to the approach in which predictors are selected if they demonstrate high levels of predictive validity with overall job performance. As Schneider et al. (1996) and others (e.g., Murphy & Shiarella, 1997; Pulakos, Borman, & Hough, 1988) have argued, this type of approach will likely enhance criterion-related validity and improve our understanding of the nature of the relationships between predictors and criteria. Viewed from this angle, synthetic validity is an approach that meets Guion (1976) and Dunnette’s (1963) call for validation strategies for scientific understanding.

Further, in most synthetic validity approaches, the validity evidence is established for each predictor against the component it is designed to measure. This process gives meaning to predictor scores in terms of the dimensions of performance (Cascio, 1998; Schneider et al., 1996) and thus, is an improvement over the will-o’-the-wisp fashion of predictor selection that Ghiselli (1959) lamented. This type of approach increases the likelihood that we are selecting the most important and relevant predictors for a job or job family (Schneider et al., 1996). Thus, the synthetic validity approach fits well with the calls for expanded and multivariate approaches to job performance (e.g., Campbell, McCloy, Oppler, & Sager, 1993; Murphy & Shiarella, 1997) and offers a potentially richer understanding of performance than other validation strategies.
Lingering Issues and Limitations of Synthetic Validity

Synthetic validity approaches are not without their limitations and unresolved issues. First of all, the results of a synthetic validity study are bound by the quality of the job analysis. Little confidence should be placed in estimated validities if the job analysis of the study was poorly conducted. Thus, synthetic validity is not a viable option in organizations that are unable or unwilling to contribute sufficient resources to job analysis initiatives. Moreover, the use of synthetic validity requires that the sample size is at least large enough to collect job analysis data on the job components. Therefore, synthetic validity will be difficult in the smallest of organizations. However, the incorporation of O*NET into synthetic validity approaches eases this limitation in some situations.

Second, the assumptions underlying a few of the methods have been critiqued. Specifically, the veracity of the "gravitational hypothesis," which is evoked by the Lawshe and Steinberg (1955) method, and the JCV model (when predicting mean test scores) has been questioned (Algera & Greuter, 1989; Trattner, 1982). Although several studies have empirically demonstrated the viability of this assumption (e.g., Sparrow, 1989; Wilk et al., 1995), it is not a guarantee that the highest ability individuals will populate the job with the highest number of critical components (e.g., the Dilbert Principle; Adams, 1996). This assumption needs to be carefully considered when using synthetic validity strategies that require it (Algera & Greuter, 1989).

The third is the lack of discriminant validity (Guion, 1998). That is, equations developed in one occupation or job can be equally applied in another. As a result, synthetic validity is more useful for selection than placement. The lack of discriminant validity could be a result of human development toward specific abilities or that performance on the job components is not specific, but general (Guion, 1998; Peterson et al., 2001). Alternatively, the lack of the discriminant validity could be a byproduct of the job analysis methods used in synthetic validity approaches. Raymark, Schmit, and Guion (1997) and Goldstein, Zedeck, and Goldstein (2002) have argued that many job analysis methods are primarily focused on the cognitive domains of work. Using the results of these job analyses to identify predictor tests can result in test batteries that primarily consist of cognitively orientated predictors. Because cognitive ability does not demonstrate discriminant validity across most jobs (Schmidt & Hunter, 1998), it is not surprising that synthetic validity equations that are made up of cognitive tests lack discriminant validity.

In addition to the issue of cognitively slanted job analyses, synthetic validity approaches have primarily focused on a narrow range of cognitive, perceptual, and psychomotor predictors (Jeanneret, 1992), which
further complicates this issue. In any case, the flaw is not fatal for any of the synthetic validity approaches, especially when components are highly correlated (Guion, 1998). Nevertheless, developing approaches that increase discriminant validity is clearly a pressing need to expand synthetic validity beyond personnel selection. As is discussed below, widening the criterion space and the range of predictors covered in synthetic validity efforts is a good step in this direction.

There are several operational details that continue to be an issue. First is the weighting of the predictors in the synthetic validity equations. There are several different weighting strategies that must be considered, and slightly different results may occur depending on the weighting strategy employed (e.g., Hoffman et al., 2000; Peterson et al., 2001). In some of the synthetic validity approaches, the weighting is based on expert judgment as opposed to statistical methods. As previously noted, there is nothing inherently inappropriate about using expert judgments (Dawes, 1979; Sawyer, 1966), especially when considering the evidence that certain experts are capable of adequately making these judgments (e.g., Peterson et al., 2001; Schmidt et al., 1983). In certain situations, these types of estimates are the only ones possible (Guion, 1998). Having said that, it is still a suboptimal weighting method (Dawes, 1979) and there are reasons to be cautious when using these judgments (e.g., Meehl, 1954). At this point, there are no clear guidelines, and as a result, these decisions will likely be driven by practicality instead of technical or statistical superiority.

Another issue is the use of tests other than those originally included in the synthetic validity study in forming a test battery. As Mossholder and Arvey (1984) note, many synthetic validity demonstrations use tests that are not commercially available, and therefore, developing a database of synthetic validity evidence that can be used by other researchers is difficult. There is some evidence that commercially available and unavailable cognitive ability tests may be exchangeable (e.g., cognitive ability tests; Hoffman & McPhail, 1998). However, research establishing the equivalence of specific tests is still needed. The use of approaches, such as those advocated by Turban, Sanders, Francis, and Osburn (1989), would be particularly helpful in establishing equivalence between predictor tests and establishing a database of synthetic validity evidence.

Recently, Harvey and colleagues (e.g., Brown & Harvey, 1996; Wagner & Harvey, 2004) have suggested that the statistical models on which some of the synthetic validity approaches are based may be inadequate. Specifically, they question the use of additive regression models (i.e., only main effects) to link the job components and mean test scores in JCV research. In Harvey and colleagues’ research, they found significant interaction effects for gender. That is, the estimated mean test scores may be impacted by the gender distribution of an occupation. Any potentially
adverse impact that may result from the use of synthetic validity equations would be of great concern to researchers and practitioners alike. However, there are many reasons for why significant gender-based interactions would be found, such as differential occupational preferences (Wagner & Harvey, 2004) or the use of predictor tests with stable gender differences such as the Myers-Briggs Type Indicator (e.g., Brown & Harvey, 1996). Much more research is needed to replicate these findings and explore the possibility of interactions with other demographic characteristics before any definitive conclusions can be made. Nevertheless, it is important to note that including interaction terms based on demographic characteristics (e.g., gender, race, age, disability status) may have legal implications. The use of these interaction terms creates differential test scores that are based on demographic characteristics. Thus, the interaction terms may be considered as an instance of within-group norming, which is prohibited under the Civil Rights Act of 1991.2

Lastly, the obscurity of the literature has been a limitation to a more widespread adoption of synthetic validity. For example, a large number of the synthetic validity references in Hoffman and McPhail (1998), Jeanneret (1992), and Mossholder and Arvey (1984) refer to technical reports and unpublished conference presentations that are not necessarily accessible to most readers. Even in this review, several of the references are drawn from conference presentations, dissertations, and technical reports. Although some of this work may eventually be published, the historical trend is that a large portion of the research is not widely disseminated. Only four Annual Review chapters mention synthetic validity (Hakel, 1986; Hough & Oswald, 2000; Schmitt & Robertson, 1990; Tenopyr & Oeltjen, 1982) and then only in passing. Moreover, many recent textbooks on personnel selection or job analysis provide little or no coverage of synthetic validity (e.g., Brannick & Levine, 2002; Schmitt & Chan, 1998; see Fleetwood & Field, 2001 for an exception). Interestingly, the topic received more coverage in earlier books (Arvey, 1979; Dunnette, 1966; Schneider, 1976). The recent increase of publications using synthetic validity will hopefully stem this trend.

Relationship with Other Validity Concepts

A central argument of this article is that synthetic validity is appropriate and useful for establishing validity evidence. In fact, synthetic validity is consistent with current conceptualizations of construct validity (e.g., Jeanneret, 1992). For example, Messick’s (1981) approach to construct

---

2I would like to thank an anonymous reviewer for this suggestion.
validity, as it is applied to jobs, requires the identification of the requirements for job performance and then the development of tests for those requirements. This process is essentially what is done in synthetic validity applications, especially JCV applications (Jeanneret, 1992). Synthetic validity is also consistent with the best practices for establishing construct validity (e.g., Society for Industrial and Organizational Psychology, 2003). Moreover, synthetic validity provides great flexibility in the possible strategies for establishing validity evidence. In fact, each synthetic validity effort reported to date is based on one or more different validation strategies. For example, the methods described by Guion (1965) and Hollenbeck and Whitener (1988) rely on a criterion-related strategy for establishing validity evidence. The J-coefficient method, on the other hand, relies on a combination of content and criterion-related strategies for establishing validity evidence.

Recent theoretical and empirical work has explicated the relationship between meta-analytic validity generalization techniques (VG) and synthetic validity (e.g., Hoffman & McPhail, 1998; Jeanneret, 1992; Morris et al., 2003). As noted above, the empirical work has found that synthetic validity and VG estimates are similar, but the synthetic validity estimates are more conservative. Again, this finding is not completely surprising given that VG estimates are typically corrected for unreliability and range restriction whereas the synthetic validity estimates are not. Yet, the difference between uncorrected VG estimates and synthetic validity estimates is very small (e.g., Hoffman & McPhail, 1998).

Conceptually, synthetic validity and VG techniques are also quite similar in several regards (Jeanneret, 1992; Mecham, 1985). First, both have the goal of establishing the generality of validity evidence across jobs or organizations. Second, both are indirect forms of validity evidence (McCormick, 1959). That is, neither is a validation study in the traditional sense of directly linking predictors and criteria. Third, both make the assumptions that when jobs have a component in common, the human attribute(s) required for performing that component will be the same across jobs and that the validity of a predictor for a particular job component is fairly constant across jobs (Hoffman & McPhail, 1998; Jeanneret, 1992; Mecham, 1985). Jeanneret (1992) has argued that there is little difference between synthetic validity (particularly the JCV approach) and VG given that they share these assumptions. Mecham (1985) has even suggested that VG is subsumed in the JCV approach to synthetic validity.

Clearly, there are operational differences between the two methods. There are also some practical and conceptual differences. Three differences are highlighted here. First is the state of the development of the technology. VG techniques are far more developed than synthetic validity techniques (Campbell, 1990). Second is the use and detail of job
analysis data. For VG purposes, Pearlman (1980) and Schmitt, Hunter, and Pearlman (1981) have argued that only a holistic job analysis is needed and that reference to the specific tasks, attributes, and behaviors is unnecessary. Specific tasks, attributes, or behaviors are necessary to establish the estimated validity of a test battery in all of the synthetic validity approaches. Third, each takes a different view of the dimensionality of job performance. In synthetic validity, performance is treated as multidimensional (Cascio, 1998). In VG, performance is typically treated in a more unidimensional manner (Murphy & Shiarella, 1997). However, it is certainly possible to minimize these differences and further integrate these techniques. For example, if a database of synthetic validity estimates for predictors and job components were developed, as is advocated by Hough (2001), VG analyses could be applied to these data to help support the inferences about the relationships between predictors and job components.

Legal Considerations

Any discussion of test validation would be incomplete without a consideration of the legal issues. To date, the legal defensibility of synthetic validity is relatively untested despite its lengthy history. Trattner (1982) and others (Hamilton, 1981; Hess, 1973) have argued that some of the synthetic validity approaches are in line with the Uniform Guidelines for Employee Selection Procedures (1978). However, Trattner (1982) has argued that some aspects of a few of the methods are not acknowledged by the Guidelines. For example, the “gravitational hypothesis” is not explicitly recognized by the Guidelines. Moreover, many approaches, such as the JCV, do not directly link test scores and job performance, which is advocated by the Guidelines. More recent treatments of validity (e.g., Messick, 1995) and the easing of some court standards are likely to have mitigated a few of the concerns identified by Trattner (1982) for some of the methods, especially the JCV. There has been only a limited number of court cases (e.g., McCloy et al. v. Willamette Industries, 2002; Taylor v. James River Corporation, 1989), involving selection procedures based on synthetic validity (i.e., JCV). Both decisions were in favor of the synthetic validity approach, but they were summary judgments and hold little legal weight.

Although future legal challenges will clarify the issue of defensibility, synthetic validity should be defensible for several reasons. First, to conduct a synthetic validation study, a comprehensive job analysis is a required. Noncomprehensive job analyses have been the legal Achilles’ heel of other validity generalization techniques (Harvey, 1991b). Second, synthetic validity studies are designed to create a test battery that will be job relevant. The predictors are chosen so that they measure the important components
of the job. The significance of relevance cannot be understated (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Society for Industrial and Organizational Psychology, 2003). Third, based on the trends of recent decisions at the time, Varca and Pattison (1993) speculated that expanding the validity frontier to strategies like synthetic validity may be less precarious than it once was.

At this point, however, little specific advice can be offered in terms of the defensibility of the various synthetic validity methods or practices. However, because all of the methods require some form of a job analysis and a clear conceptualization of performance, synthetic validity may be more defensible than other validity generalization techniques. In addition, the recent update of the Principles for the Validation and Use of Personnel Selection Procedures by the Society for Industrial and Organizational Psychology (2003) explicitly discusses synthetic validity as a means to establish validity and validity generalization, thus bolstering its status as a possibly defensible strategy. Of course, the quality of the synthetic validity study, the appropriateness of the chosen strategy, and the nature of inferences one attempts to make will determine the ultimate defensibility.

Future Directions

Despite the impressive advances that have been made, the full potential of synthetic validity has yet to be realized (Hoffman et al., 2000; Mossholder & Arvey, 1984). There are many potential directions for synthetic validity practice and research. In terms of methodological and analytic strategies, McCloy’s (1994, 2001) integration of multilevel modeling and synthetic validity indicates several potential applications and developments for synthetic validity. Specifically, multilevel modeling could be used to bridge the synthetic validity and dynamic criteria literatures. For example, one could add time as a third level of analysis to the equations presented by McCloy (1994, 2001). This analysis would indicate how the effects of job components on the predictor and criteria relationship vary over time. These types of analyses may also increase the usefulness of synthetic studies when jobs are changing rapidly in that the data could be used to create time-specific test batteries. Thus, for a given job, there could be multiple test batteries and synthetic validity equations at different points in time.

Interestingly, this method could test the veracity of some of the underlying assumptions that some synthetic validity methods are based on. For example, this method could test the assumption that the validity of a test for predicting component performance is similar across jobs. One could add a job level of analysis to the equations presented by McCloy.
(1994, 2001) to determine if the validity is actually similar across jobs. This method potentially offers a major conceptual and methodological advance in synthetic validity research.

Recently, a considerable amount of research has been devoted to comparing VG estimates and synthetic validity estimates. Much less work has been directed at comparing the validity estimates from different synthetic validity approaches with each other and then with VG estimates. The initial evidence suggests that it may be the case that some synthetic validity methods produce estimates that are closer than others to VG estimates. For instance, Jago (1995) found that the JCV approach produced estimates that were closer to VG estimates than the J-coefficient method. However, more comparative research is needed.

A second potential application is the extension of synthetic validity to a wider range of predictors. Synthetic validity studies have primarily relied on the measurements of cognitive, perceptual, and psychomotor abilities (Jeanneret, 1992). Although there are exceptions (e.g., situational judgment tests, Johnson, Carter, et al., 2001; physical ability tests, Hoffman, 1999; personality inventories, D'Egidio, 2001), the focus overall has been narrow and a limitation of synthetic validity. With the development of reliable measures of personality and vocational interests, these constructs could easily be further integrated into synthetic validity efforts. In addition, assessment center ratings or conditional reasoning tests could be other avenues to expand the range of predictors used. Schneider et al. (1996) discuss one way in which the predictor domain could be expanded using personality and Hollenbeck and Whitener's (1988) synthetic validity approach. Brown and Harvey (1996), D'Egidio (2001), and Johnson, Carter, et al. (2001) present some initial empirical evidence supporting these types of extensions and Hough (2001) provides the conceptual framework to support these efforts. Additional research is certainly needed to further expand synthetic validity on the predictor side.

In the same vein, expanding the range of criteria is another direction that future work could take. To date, the majority of the research on synthetic validity has focused on job performance ratings. As is true for the predictors, this narrow focus has been a limitation. Synthetic validity techniques could be used with other organizationally relevant criteria such as training performance, work samples, counterproductive work behaviors, job attitudes, or withdrawal behaviors. Drawing on the meta-analytic work of Nathan and Alexander (1988) and the research of Pulakos et al. (1988), it is feasible that the use of criteria other than performance ratings may result in different levels of predicted validity. Synthetic validity also has much to offer to the calls for expanding the criterion domain (e.g., Borman & Motowidlo, 1993). Incorporating contextual performance criteria is a natural direction for synthetic validity research to take. Johnson, Carter, et al. (2001) present some initial findings that are supportive of
incorporating contextual performance criteria. Additional research on these types of criteria and other criteria is needed.

Further work in both the predictor and criteria domains would be in line with Hough's (2001; Hough & Ones, 2001) call for the development of a database of the relationships between predictors and job components for use with synthetic validity approaches. This database could be used to create predictive equations for specific situations. Hough and colleagues suggest that such a database could be developed from meta-analyses of primary studies of the relationships between predictors and job components. Moreover, these efforts would target narrower (i.e., less compound) predictor constructs, as well as more specific criteria. The potential result is even more tailored and precise synthetic validity equations than has been possible in previous synthetic validity efforts.

According to Hough (2001), this database will allow I-O practitioners and researchers to use synthetic validity to better meet the changing prediction needs of organizations. This type of approach could also address two issues surrounding synthetic validity. First, it would reduce the reliance on expert judgment in the estimates of the relationships between predictors and job components for some of the methods. Second, it is likely to help improve the discriminant validity of synthetic validity equations. As the equations become more narrowly focused on a particular job or situation and include a wider range of predictors, it is less likely that the equations will be equally predictive for other jobs and situations.

The usefulness of synthetic validity methods has been demonstrated primarily on entry-level jobs in business and the military. Even then, the majority of the research has utilized clerical jobs. Future applications are needed that extend synthetic validity applications to managerial, professional, and executive jobs. Little is known about the usefulness of synthetic validity with these types of jobs. Logically, synthetic validity may be very useful with these jobs because of the difficulty of obtaining large sample sizes. Although there will be unique difficulties with many of these types of jobs (e.g., conducting job analyses), the potential to contribute to the increased prediction accuracy for these types of jobs cannot be overlooked.

The current changes in the structure and organization of work have several implications for synthetic validity. Up to this point, synthetic validity methods have been applied to individual-level selection. However, this method could be extended to team selection. As the use of teams continues to grow, the application of synthetic validity to team selection may provide a fruitful area for additional work. If the current trend of dejobbing continues (Bridges, 1994; Cascio, 1995), the usefulness of synthetic validity will grow dramatically. Because it uses components of jobs, it can be applied when no jobs exist, when a job is redesigned, or when jobs are rapidly changing. Lastly, synthetic validity may be extended beyond the organizational selection domain. For example, it may be possible to apply
synthetic validity principles to training program validation, educational selection, or educational and professional certification.

Conclusions

The history of synthetic validity has been marked with periods of great interest and other periods of great neglect. This state of affairs is unfortunate given the practical and conceptual advantages that synthetic validity can offer. Synthetic validity is currently the only validation strategy that can be applied to the vast majority of public, private, military, and government organizations. Fortunately, interest over the past 10 years has dramatically increased and many substantial new developments have been offered. Much has been accomplished and many interesting questions remain. Mossholder and Arvey (1984) had hoped that additional developments would transform synthetic validity from a promising technique to a well-developed and prominent technique. Twenty years later, it appears that their wish may be finally coming true.

REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


