RESEARCH REPORTS

The Importance of Distinguishing Between Constructs and Methods When Comparing Predictors in Personnel Selection Research and Practice

Winfred Arthur Jr. and Anton J. Villado
Texas A&M University

The authors highlight the importance and discuss the criticality of distinguishing between constructs and methods when comparing predictors. They note that comparisons of constructs and methods in comparative evaluations of predictors result in outcomes that are theoretically and practically misleading. The objective in this article is (a) to highlight the importance of distinguishing between predictor constructs and predictor methods when comparing predictors in personnel selection research and practice and (b) to discuss the theoretical and practical implications of the failure to draw and maintain this distinction. Our primary thesis is that streams of personnel psychology research being used to frame this discussion. Researchers, editors, reviewers, educators, and consumers of research are urged to carefully consider the extent to which the construct–method distinction is made and maintained in their own research and that of others, especially when predictors are being compared. It is hoped that this discussion will reorient researchers and practitioners toward a more construct-oriented approach that is aligned with a scientific emphasis in personnel selection research and practice.

Keywords: comparing predictors, predictor constructs, predictor methods, subgroup differences, applicant reactions

The objective in this article is (a) to highlight the importance of recognizing and maintaining the distinction between predictor constructs and predictor methods when comparing predictors in personnel selection research and practice and (b) to discuss the theoretical and practical implications of the failure to draw and maintain this distinction. Our primary thesis is that streams of research or applied effort that involve the comparison of predictors must recognize and maintain the distinction between constructs and methods. Unless this is done, the resultant comparisons are problematic, because they end up being comparisons of predictor constructs to predictor methods. In addition, when predictor constructs and predictor methods are confounded, it is impossible to determine whether the observed effects from the comparisons are due to what is being measured or to how it is being measured.

The Predictor Construct–Predictor Method Distinction

Among personnel psychologists, there has been a long-standing interest in the development of predictors for use by those making a variety of employment decisions. Accordingly, comparative evaluations of predictors are quite common in the literature (e.g., Ghiselli, 1973; Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984), culminating, for instance, in Schmidt and Hunter’s (1998) meta-analysis, which summarized 85 years of research findings on “the validity and utility of selection methods in personnel psychology” (p. 262). However, a closer reading of this sort of comparison clearly indicates that the studies are comparisons of predictor constructs and predictor methods (e.g., see the first columns of Schmidt and Hunter’s Table 1 and Table 2).1

A predictor is a specific behavioral domain, information about which is sampled via a specific method. Thus, depending on one’s focus, predictors can be represented in terms of what they measure and of how they measure what they are designed to measure. The behavioral domain of predictors (i.e., what they measure) can be delineated by theories of psychological constructs (e.g., knowledge, skills, and abilities), theories of job situations/demands, or even some combination of the two. In this article, we use the term predictor construct (or construct) to refer to the behavioral domain being sampled, regardless of how that domain is specified. On the other hand, we use the term predictor method (or method) to refer to the specific process or techniques by which domain-relevant behavioral information is elicited, collected, and subsequently used to make inferences (Arthur, Day, McNelly, & Edens, 2003; Arthur & Doverspike, 2005; Arthur, Edwards, & Barrett, 2002; Arthur & Doverspike, 2005; Arthur, Edwards, & Barrett, 2002;

---

1 Although we cite and review specific studies in this article as examples of the predictor construct–predictor method confound in the domains of interest, no indictment or critique of these studies is intended. Instead, these examples are intended to highlight and document what we consider to be a rather pervasive occurrence, even in top-tier scientific journals on industrial/organizational psychology and human resource management.
J. P. Campbell, 1990). Given this distinction, an emphasis on predictor constructs focuses on the behavioral domain of the predictor, whereas an emphasis on predictor methods focuses on the method of obtaining information on the behavioral domain. Predictor constructs may include or take the form of psychological constructs and variables, such as general mental ability, conscientiousness, psychomotor ability, and perceptual speed. They can also take the form of situational or job-content-based behaviors, such as word processing or troubleshooting an F-16 jet engine. In contrast, predictor methods may take the form of interviews, paper-and-pencil tests, and computer-administered, video-based, or simulation-based modes of assessment.

In this article, we demonstrate the criticality and the theoretical and practical implications of the predictor construct–predictor method distinction, using three important streams of personnel psychology research to frame this discussion. These three streams of research are (a) investigations of predictor-criterion-related and incremental validity, (b) techniques for reducing predictor subgroup differences, and (c) applicant reactions to predictors in selection.

From a theoretical perspective, the construct–method distinction is important, because it allows the isolation of variance due to predictor constructs from the variance due to predictor methods (D. T. Campbell & Fiske, 1959). This isolation of variance permits comparisons that are theoretically and conceptually interpretable and meaningful. From a research design and methodology perspective, Campbell and Fiske’s multitrait–multimethod framework, as illustrated in Figure 1, suggests that comparisons of predictor constructs should be column comparisons (e.g., A vs. E vs. I vs. M), such that the predictor method (i.e., M1) is held constant (e.g., high-structure interview) while the predictor constructs (i.e., C1, C2, C3, and C4) are varied (e.g., general mental ability vs. conscientiousness vs. agreeableness vs. locus of control). In contrast, when the focus is on predictor methods, the appropriate comparison should hold the predictor construct constant (e.g., general mental ability) and vary the predictor methods (e.g., paper-and-pencil test vs. interview vs. performance). That is, the most appropriate comparative evaluation of predictor methods is one that is done within constructs. In Figure 1, the most appropriate comparisons would be those conducted within rows (e.g., those comparing Cells A vs. B vs. C vs. D or Cells I vs. J vs. K vs. L) but not those across rows and columns, which compare constructs and methods (e.g., those comparing Cells A vs. F vs. K or Cells B vs. E vs. P). The latter comparisons confound predictor constructs and predictor methods and make it impossible to disentangle the sources of observed effects. Hence, these sorts of comparisons (e.g., comparative levels of subgroup differences displayed by general mental ability and assessment centers) make research results difficult, if not impossible, to interpret (see, e.g., Goldstein, Yusko, Braverman, Smith, & Chung, 1998; Hoffman & Thornton, 1997).

The importance of the predictor construct–predictor method distinction is embedded in the centrality of constructs to psychology as a scientific discipline (Murphy & Davidshofer, 2001; Nunnally & Bernstein, 1994). This fact is probably best highlighted by Binning and Barrett’s (1989) presentation of the unitarian view of validity (see also Landy, 1986). Figure 2 represents an adaptation of Binning and Barrett’s Figures 2 and 3. Consequently, our description of this framework borrows extensively from their work. The interested reader should refer to the original article for a detailed treatment of these issues.

As noted by Binning and Barrett (1989), the process of validating selection procedures can be viewed as a special case of hypothesis testing and scientific theory building (pp. 478–479). Figure 2 therefore illustrates a logical system of personnel selection inferences that represent the conceptual linkages between the predictor construct domain and the job performance domain, both of which are representations of the underlying construct domain. Thus, in Figure 2, Inference 1 represents the proposition that the predictor construct domain is theoretically or conceptually associated with the job performance domain by virtue of the fact that they are both representations of the underlying construct domain.

The job performance domain, which is defined in terms of job behaviors or outcomes, can be developed through job analysis or established via organizational policy. Consequently, Inference 1 is established via job analysis processes that are intended to identify the predictor constructs deemed requisite for the successful performance of the specified job (or performance) behaviors in question. As a result, the following logical inferences are portrayed in Figure 2:

Inference 2 implies that the predictor measure is an adequate sample from [the] psychological construct domain; Inference 3 implies that the criterion measure is an adequate sample from the performance domain; Inference 4 implies that predictor measurements relate to criterion measurements; and Inference 5 implies that the predictor measure is related to the performance domain. (Binning & Barrett, 1989, p. 480)

In summary, this logical system of inferences, which is derived from the unitarian view of validity, highlights the fact that the validation of specified predictors cannot be divorced from a discussion of what the predictors are designed to measure. This requisite exists because the predictor and criterion measures ultimately are theoretically and conceptually linked to an underlying construct domain.

In a similar vein, this line of reasoning is consistent with the view that, from a theoretical perspective, individual differences are conceptualized in terms of predictor constructs. So, in personnel selection, for example, the theoretical determinants of job performance and other valued outcomes are ultimately conceptualized in terms of predictor constructs and not predictor methods. This is not

<table>
<thead>
<tr>
<th>CONSTRUCT</th>
<th>METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
</tr>
<tr>
<td>C1</td>
<td>A</td>
</tr>
<tr>
<td>C2</td>
<td>E</td>
</tr>
<tr>
<td>C3</td>
<td>I</td>
</tr>
<tr>
<td>C4</td>
<td>M</td>
</tr>
</tbody>
</table>

Figure 1. Multitrait–multimethod matrix for comparing predictors on outcomes, such as criterion-related and incremental validity, test-taker reactions, and levels of subgroup differences.

---

2 We acknowledge the input of an anonymous reviewer, who assisted us in arriving at this more precise and succinct delineation.
to imply that a study cannot focus on predictor methods. However, we emphasize that to theoretically explain why scores obtained from even a predictor method (e.g., the employment interview) are related to the outcomes of interest, one almost invariably has to invoke a predictor construct explanation (e.g., Huffcutt, Conway, Roth, & Stone, 2001). As noted by Schmidt and Hunter (1998), the goals of personnel psychology include more than a delineation of relationships that are practically useful in selecting employees. Hence, in recent years, the focus in personnel psychology has turned to the development of theories of the causes of job performance [with] the objective of understanding the psychological processes underlying and determining job performance. (p. 271)

An implication of the construct–method distinction is that, all things being equal, any specified predictor method can be designed to assess a wide array of predictor constructs. Of course, methods may differ in the ease with which they can be designed to measure some constructs. Conversely, a given predictor construct can be measured with a wide array of predictor methods. This orientation is illustrated in Figure 1. So, for example, the interview has been used to assess constructs such as interpersonal skills, general mental ability (Huffcutt et al., 2001), and even personality variables (Van Iddekinge, Raymark, & Roth, 2005). In turn, general mental ability has been measured with methods such as performance tests, paper-and-pencil tests (e.g., Raven’s Advanced Progressive Matrices; Raven, Raven, & Court, 1994), and even biological and physiological measures (see Matarazzo, 1992). The Wechsler Adult Intelligence Scale (Wechsler, 1972) is an example of an ability test in which the test stimuli are presented orally and the test taker responds to some items by performing a task and to others by responding orally. Likewise, the ability to troubleshoot an F-16 jet engine can be measured via hands-on performance with an actual engine, a computer simulation, or an oral walk-through protocol (Hedge & Teachout, 1992).

Criterion-Related Validity and Incremental Validity

Predictors are frequently compared in terms of their relative criterion-related validity and incremental validity. Yet, a closer reading of this literature indicates that most of these studies are comparisons of predictor constructs to predictor methods. Thus, the criterion-related validities of general mental ability and conscientiousness have been compared to those of interviews (Campion, Campion, & Hudson, 1994; Cortina, Goldstein, Payne, Davidson, & Gilliland, 2000; Dipboye, Gaugler, Hayes, & Parker, 2001); assessment centers (Dayan, Kasten, & Fox, 2002; Goffin, Rothenstein, & Johnston, 1996); and a wide array of other methods (e.g., Callinan & Robertson, 2000; Hermelin & Robertson, 2001; Mount, Witt, & Barrick, 2000; Villanova, Bernardin, Johnson, & Dahmus, 1994). Consequently, from a theoretical and conceptual perspective, it is unclear what a comparison of a predictor construct and a predictor method represents. This issue is exacerbated when the comparisons involve overall assessment center ratings (OARs; e.g., Dayan et al., 2002; Goffin et al., 1996), because the OAR aggregates multiple dimension (i.e., predictor construct) scores to arrive at a single (predictor method) score. Consequently, the use of the OAR results in a loss of information at the predictor-construct level (Arthur et al., 2003).

From both a theoretical and a practical perspective, there are a number of potential critiques of the views espoused in this article that are worth addressing. The first is Schmidt and Hunter’s (1998) argument that the failure of meta-analyses to support the situational specificity hypothesis “indicates that such methods as interviews, assessment centers, and biodata measures do not vary much from application to application in the constructs they measure” (p. 271). Furthermore, whereas Schmidt and Hunter observe that whereas “we do not know exactly what combination of constructs is measured by methods such as the assessment center, the interview, and biodata. . . . whatever those combinations are, they do not appear to vary much from one application [study] to another” (p. 271). The conclusion of this critique, then, is that there may be little or no reason to be concerned about method-level data. We have multiple responses to this critique. First, the issue is not the meaningfulness of method-level data but the comparison of predictor methods to predictor constructs, which takes the form of making comparisons across rows and columns in Figure 1. Our position is that when the focus is on predictor methods, the appropriate comparisons should be within rows, as shown in Figure 1.

Second, a closer reading of meta-analyses of predictor methods indicates that the overall analysis rarely explains all the variability in the corrected validities (e.g., Gaugler, Rosenthal, Thornton, & Benton, 1987; Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Roth, Bobko, & McFarland, 2005). Instead, the overall analysis is typically followed by a search for moderators, which are often method-level rather than construct-level moderators. We would argue that, unless the constructs are held constant in the initial overall analysis, a search for construct-level moderators (examples of which include Arthur et al., 2003, and Huffcutt et al., 2001) is a more meaningful approach than is a search for method-level moderators. This response is in line with the Principles for the Validation and Use of Personnel Selection Procedures (Society for Industrial and Organizational Psychology [SIOP], 2003), which has clearly and explicitly stated that one

3 It is noteworthy that a defining characteristic of assessment centers is the use of multiple exercises (methods) to measure multiple dimensions. The exercise scores are then aggregated to obtain the dimension scores.
needs to consider constructs when conducting a meta-analysis of predictors. Furthermore, the Principles has observed that “when studies are cumulated on the basis of common methods (e.g., interviews, biodata) instead of constructs, a different set of interpretational difficulties arise” (p. 30), namely, to what extent do the predictor methods measure the same predictor construct?

The problem alluded to here is reflected in the assessment center meta-analyses literature. Because OARs are aggregates of constructs, comparisons of assessment centers at the level of OARs are almost guaranteed to result in divergent findings, because, after all, the comparisons involve amalgamations of different constructs. That is, no two OARs are necessarily going to be the same unless the dimensions that make up the OARs are held constant. So, from this perspective, it is not surprising that meta-analytic estimates of the criterion-related validity of assessment centers, operationalized as OARs, have ranged widely, with estimates of .43 (Hunter & Hunter, 1984), .41 (Schmitt et al., 1984), .37 (Gaugler et al., 1987), .26 (Hardison & Sackett, 2004), and .22 (Aamodt, 2004). Although these differences could be due to differences in methodology, inclusion criteria, and historical changes in the quality of assessment centers, we submit that a major potential reason for the range and divergence in assessment center findings may be the pervasive focus on OARs in these meta-analyses.

A second, more applied critique of the construct–method distinction is that the confound (i.e., the failure to disaggregate predictor constructs and predictor methods) is not a concern, because this is the way personnel measures are conceptualized and used in the field (Schmidt & Hunter, 1998). So, for instance, if the interview yields incremental validity over general mental ability, or if a multiple-predictor selection system that includes general mental ability, an interview, and assessment center yields a high multiple $R$, this result has a lot of informational value for the practitioner’s organization. Consequently, why should one care about what constructs are being measured by the interview and assessment center?

Although we acknowledge that this approach may indeed reflect how these personnel measures are generally used in the field, the constructs being measured should be of concern. Whereas a study like the one described above may have informational value for the extant organization, unless one assumes that the validity of predictor methods is construct invariant—an assumption that is likely to be unfounded (see, e.g., the assessment center and interview meta-analyses of Arthur et al., 2003, and Huffcutt et al., 2001, respectively)—the study has no informational value for other practitioners and organizations that may want to use the results as the basis for the development of their own selection test batteries. The likelihood of obtaining the same validities is likely to be hit or miss, with the outcome determined primarily by the extent to which the constructs measured by their own interviews and assessment centers approximate those measured by the original study. This argument, of course, brings us back to the central and critical role of constructs in the comparative evaluation of predictors. That is, for the community of practitioners, the informational value of validation studies conducted by others will be much higher if information is provided about the constructs that were assessed by the predictor batteries (see, e.g., LaHuis, Martin, & Avis, 2005, and Liewens & Sackett, 2006).

The preceding discussion provides an opportunity to highlight the issue of espoused constructs versus actual constructs. The issue here is one of construct validity, and, in the context of meta-analyses and of practitioners reporting research results, it is important to stress that labeling data as reflections of a particular construct does not mean that the construct is being assessed. This fact is probably best reflected in assessment center research and practice, a domain in which statements about what exercises measure (e.g., sensitivity, self-direction, inspiring trust, seasoned judgment, personal breadth; see Table 2 of Arthur et al., 2003, for additional examples) are typically by self-proclamation and in which construct-related (i.e., discriminant and convergent) validity evidence in support of these assertions (Woehr & Arthur, 2003) is rarely presented.

**Techniques for Reducing Subgroup Differences**

As with the research comparing the criterion-related and incremental validities of predictors, research on techniques for reducing subgroup differences also suffers from the predictor construct–predictor method confound. Industrial/organizational psychologists have long provided advice to organizations on how to reduce subgroup differences and adverse impact (Guion, 1998). For instance, in light of the magnitude of subgroup differences associated with general mental ability, the use of nonability predictor constructs in personnel selection has been considered as a way to reduce these differences (Hogan, Hogan, & Roberts, 1996; Schmitt, Clause, & Pulakos, 1996). A variation of this construct-change approach has been to combine general mental ability with other predictors (Sackett & Ellingson, 1997; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). From the perspective of the construct–method distinction, these predictors have sometimes been constructs (e.g., personality variables, integrity; Avis, Kudisch, & Fortunato, 2002; Bing, Whanger, Davison, & VanHook, 2004; Ryan, Ployhart, & Friedel, 1998). However, more often than not, they have been predictor methods (e.g., structured interviews, performance tests; Bobko, Roth, & Potosky, 1999; Goldstein et al., 1998; Sackett & Ellingson, 1997; Schmitt et al., 1997).

The problems associated with the construct-and-method approach are the same as those associated with the method-change approach. Specifically, as an approach to reducing subgroup differences, the method-change approach has included the use of performance tests (Chan & Schmitt, 1997), assessment centers (Goldstein et al., 1998), situational judgment tests (Nguyen, Mcdaniel, & Whetzel, 2005), audiotape-recorded interviews (Raye, Curtis, Snyder, & Satterwhite, 2005), and video-based tests (Chan & Schmitt, 1997; Schmitt & Mills, 2001; Weekley & Jones, 1999). This stream of research has resulted in widely professed conclusions, one being that adverse impact is less of a problem with assessment centers and work samples than it is with general mental ability (Cascio & Aguinis, 2005, p. 372; Hoffman & Thornton, 1997; cf. Bobko, Roth, & Buster, 2005).

A problem with the method-change and construct-and-method combination approaches is that, because these approaches fail to disaggregate predictor constructs and predictor methods, it is unclear whether the observed reductions in subgroup differences are due to changes in the constructs being measured or to the methods of assessment. So, for example, in the case of assessment centers, because the methods and constructs are confounded, it is unclear whether the reduction in subgroup differences is due to the assessment method used or to the constructs measured (Arthur et al.,
In short, any method of assessment can display high or low levels of subgroup differences; it depends on the construct or constructs being measured. From a practical perspective, to clearly delineate which predictor will result in a lower level of subgroup differences, we need to know both the possible methods of assessment (e.g., work samples, interviews, assessment centers) and the constructs assessed by those methods.

Applicant Reactions to Selection Tests

Anderson (2004) noted that the focus of personnel selection research on the applicant perspective, in contrast to the organizational perspective, is relatively recent and was long overdue. This stream of research investigates applicant reactions to selection systems, processes, methods, and decisions and the relationships of these reactions to outcomes, such as perceptions of fairness, face validity, test-taking motivation, test performance, and self-withdrawal from the selection process (Anderson, 2004; Chan & Schmitt, 2004; Ryan & Ployhart, 2000). Consequently, a common paradigm is for studies to compare applicants’ reactions to different predictors (e.g., Hausknecht, Day, & Thomas, 2004; Lievens, De Corte, & Bryssee, 2003). However, as with the comparative criterion-related validity studies, a closer reading indicates that these comparisons are an amalgam of predictor constructs and predictor methods (e.g., Kravitz, Stinson, & Chavez, 1996; Macan, Avedon, Pease, & Smith, 1994; Moscoso & Salgado, 2004; Rynes & Comerley, 1993; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993).

This confound has important theoretical and practical implications. If the goal in this stream of research is to identify and understand factors that influence applicant perceptions of test fairness, face validity, and test-taking motivation that subsequently could influence test performance and self-withdrawal from the selection process, it would seem critical to be able to determine whether the observed applicant reactions are due to the method of assessment or to what is being measured. Thus, whereas the predictor construct may be determined by the job analysis and other systematic processes, in contrast, personnel researchers and practitioners have a wide range of methods to operationalize the specified job-relevant constructs, once these constructs have been identified.

How Should Comparative Evaluations of Predictors Be Conducted?

Regardless of the research domain, attending to the predictor construct–predictor method distinction entails the same underlying principle: Research ought to be conducted in a manner that recognizes constructs and methods as two factors and should strive to implement studies that either do not confound the two or that provide information on what constructs are being measured by the methods being studied. We believe this critical point has been overlooked by many researchers in various research domains, a fact that limits the import of their findings. Despite the prevalence of studies that ignore the distinction between predictor constructs and predictor methods, exemplars that demonstrate the importance of recognizing this distinction and incorporating it into the study design exist in each of the three domains discussed. For example, Lievens and Sackett (2006) compared the criterion-related validity of a video-based and a written situational judgment test. In doing so, they held the predictor constructs assessed by the two situational judgment tests constant, which allowed them to convincingly conclude that the video-based situational judgment test produced significantly higher criterion-related validity estimates than did the written format. However, had Lievens and Sackett concomitantly varied the constructs assessed by these two test method formats, they would have found it impossible to determine whether the higher criterion-related validity observed for the video format was indeed due to the video format or to the accompanying change in what was being measured.

A 2007 study by Edwards and Arthur is another example that highlights the importance of distinguishing between predictor constructs and predictor methods. Specifically, Edwards and Arthur compared the level of subgroup differences associated with two types of paper-and-pencil test formats, constructed-response and multiple-choice, for a measure of scholastic achievement. Again, by holding the predictor construct constant and varying the predictor method, they were able to show that the 39% reduction in subgroup differences observed for the constructed-response format was due to the change in predictor method.

Finally, as an example of a study that compared test-takers’ reactions to different predictors, Bauer, Truxillo, Paronto, Weekley, and Campion (2004) investigated student reactions to one of three types of screening techniques: a face-to-face interview, a telephone interview, or an interactive voice–response interview. As with the preceding examples, because the predictor construct was identical across the three techniques, it was possible to conclude that the observed differences in applicant reactions were attributable to the specific methods.

Summary and Conclusions

There are instances in a field or discipline’s development when certain knowledge and practices are so broadly and commonly accepted that they achieve the status of received doctrines. However, although received doctrines may have their place, we sometimes run the risk of accepting them too readily or overlooking some basic conceptual flaws that may be associated with them. We posit that the comparative evaluation of predictors that entails the comparison of predictor methods to predictor constructs has achieved this status. Consequently, the present article identifies and highlights the criticality of the construct–method distinction in the comparative evaluation of predictors by using three streams of personnel psychology research as illustrative examples. Our goal is to sensitize both researchers and practitioners to the theoretical and conceptual threats and pitfalls associated with this practice, which seriously restricts the value of research information in terms of interpretation of the results of primary studies and resultant meta-analyses as well (SIOP, 2003).

We hope that highlighting these issues will engender a discussion culminating in programmatic research that will eventually allow us to fill in construct and method matrices, such as the one illustrated in Figure 1. It is also worth noting that a focus on a construct-oriented approach does not have to be at the expense of an interest in applied research. To the contrary, such a framework has enormous applied utility and implications. For instance, it provides guidance on the choice of and match between predictor constructs and predictor methods, given specified boundary con-
ditions. Thus, we acknowledge the distinction between scientific and applied utility and also recognize that, although different, they are both of value. However, we posit that, as a scientific psychological discipline, applied utility would be subsumed under scientific utility and that, in fact, a focus on the latter would facilitate the former. To paraphrase, “If it is not good science, it is not good practice.”

A focus on the predictor construct–predictor method distinction can also serve as an impetus for future research. For instance, one could conceptualize and delineate several dimensions of predictor methodology, including a distinction between the method used to present the test stimulus material and the method used by the test taker to present the response and whether the response requires the recognition or construction of a response. Likewise, scoring can be delineated as another dimension of methodology (Huffcutt & Arthur, 1994). Thus, this facet of methodology could entail the degree of subjectivity or judgment involved in arriving at the predictor scores. Although these distinctions may be too fine in some cases, they nevertheless allow us, for instance, to draw a distinction between situational judgment tests, in which the stimuli are presented by video and participants respond either orally or via paper and pencil. Likewise, they allow us to classify interviews as a method in which both the stimuli and responses are presented via an oral medium, with the standardization of the scoring of responses ranging from a global assessment to the use of preestablished answers to evaluate each individual response (see, e.g., Huffcutt & Arthur, 1994, Figure 1, p. 187).

If ignored, these dimensions of predictor methodology can serve as a potential source of construct-irrelevant variance (Messick, 1995). On the other hand, if incorporated into the generation of research questions and hypotheses and subsequent research designs, they are a potentially rich source of meaningful theoretical and practical research investigations. For example, studies that have provided interpretable data in which alternative predictor methods were examined by holding the predictor construct constant across methods include Arthur et al. (2002); Chan and Schmitt (1997); Edwards and Arthur (2007); Lievens and Sackett (2006); Richman-Hirsch, Olson-Buchanan, and Drasgow (2000); and Schmitt and Mills (2001).

There are additional facets of the predictor construct–predictor method issue that we have not addressed. However, we think they are worthy of mention and would like to introduce them to the reader as well. First, we recognize that the predictor method can influence the predictor construct being measured. That is, each predictor method potentially has its own source of construct-irrelevant variance (Messick, 1995). For instance, a predictor method that has poor face validity may influence test-taking behavior in a manner that alters the intended predictor construct that is being measured (see Chan & Schmitt, 1997, for a discussion of some of these issues). In spite of this, we maintain that the predictor construct and predictor method should be and are conceptually separate and that predictor methods should not be compared to predictor constructs.

Second, it is our observation that the failure to maintain the distinction between predictor constructs and predictor methods has resulted in some situations in which there is serious uncertainty as to whether a specified predictor is a construct or a method. Situational judgment tests are a good example of this, and several recent studies have tried to address the confusion and lack of clarity in the extant literature as to whether the situational judgment test is a construct or a method (Schmitt & Chan, 2006). Specifically, Schmitt and Chan’s goal was to ascertain whether the situational judgment test is a construct or a method. As a tentative answer, they proposed “that SJTs, like the interview, be construed as a method of testing that can be used to measure different constructs but that the method places constraints on the range of constructs measured” (p. 149). This proposal is consistent with our preceding point: that although different methods can theoretically be designed to measure any construct, in practice, they vary in the case with which they can be designed to assess some constructs. Thus, in practice (although not in theory), some cells in Figure 1 might be empty (e.g., those showing use of a performance test to measure social desirability or of a paper-and-pencil test to measure upper body strength).

In sum, Schmitt and Chan (2006) posited that the primary dominant constructs measured by situational judgment tests are “adaptability constructs,” which can be represented by the global construct of “practical intelligence” or a “general judgment factor” (p. 149). It is interesting to note that these construct labels appear quite frequently in the assessment center literature (see, e.g., Arthur et al., 2003, Table 2). So, whereas we consider an attempt to resolve this particular issue to be beyond the scope of this article, our view on this issue is similar to that of Schmitt and Chan: Situational judgment tests are methods “that can be used to assess different constructs [which are] probably conceptually distinct from established constructs such as cognitive ability and personality traits” (p. 148). Indeed, we carry this thinking one step further by submitting that situational interviews and other scenario-based tests are all best conceptualized as situational judgment tests that just happen to differ on additional method dimensions, such as video-based, oral, or paper-and-pencil administration.

In conclusion, we have presented a host of reasons, coupled with illustrative examples, for why it is important to recognize and maintain the distinction between constructs and methods when comparing predictors. We conclude by urging researchers, editors, reviewers, educators, and consumers of research to carefully consider the extent to which the distinction between predictor constructs and predictor methods is made and maintained in their own research and that of others, especially when different predictors are being compared. Finally, we hope this discussion reorients researchers and practitioners toward a more construct-oriented approach that is aligned with a scientific emphasis in personnel selection research and practice.

References


Received January 24, 2006
Revision received June 29, 2007
Accepted July 9, 2007