Projective testing: Historical foundations and uses for human resources management☆

Nathan T. Carter,⁎, Michael A. Daniels, Michael J. Zickar

University of Georgia, Department of Psychology, 323 Psychology Building, Athens, GA 30602, United States
Bowling Green State University, Department of Psychology, Bowling Green, OH 43403, United States

1. Introduction

The use of projective psychological measures has intrigued scientists and even the general public for nearly a century. In this special issue, the use of these types of measures in application to human resource management (HRM) science is explored. We believe that the examination of projective measurement’s history in its field of birth (psychology) as well as organizational science is paramount to developing an informed, well-rounded discussion that can help researchers avoid the foibles of the past. Here we provide the historical and conceptual foundations of different projective measures, and review their use in areas relevant to HRM. Based on our review, we suggest that the evident trend of slowing research in the use of many projective measures for HRM applications is not warranted given the empirical record. We note two major barriers to the use of these measures in HRM that we believe contributed to the dearth of recent research in the area.

First, we propose there is a great need for HRM-centric literature on projectives. Historically, most projective measures are rooted in theories and perspectives seated in the context of abnormal attributes. As will be noted, this fact has serious legal and ethical implications for HRM practitioners. Further, the lack of an HRM-centric literature has resulted in measures that do not always address the practical administration and scoring issues specific to HRM. We therefore advocate the development of an HRM-centric literature that: a) uses “normal” theories of personal attributes for projective test development; and b) considers the practical concerns of HRM practitioners as a central concern for researchers in the area.

Secondly, we suggest that subgroups of researchers investigating projectives within and outside of HRM realms have historically become mired in what we call philosophical deadlock, wherein the evidence provided by one group does not address the concerns of another and vice versa. We humbly provide some suggestions and hint at future directions for projective measurement in HRM.

© 2012 Elsevier Inc. All rights reserved.

Keywords: Projective testing, Personality testing, Applied psychology, History
the concerns of another, and vice versa. Based on these emergent themes in our historical review, we humbly suggest three guidelines for researchers to consider in the interest of avoiding such deadlock in the future. Namely, we argue that: a) projectives’ correlations with self-report measures are not always the “gold standard” for validity; b) different measurement methods are not always competing for the construct; and c) the focus of questions of validity should be focused on whether the construct causes variation in test scores.

We organize the following historical review according to the taxonomy of projective measures put forward by Lindzey (1959) and reiterated by Lilienfield, Wood, and Garb (2000). Table 1 outlines several clinical and HRM-oriented projective tests that fit these different categories, which are as follows: a) Association; b) Construction; c) Completion; d) Arrangement; and e) Expression. Only one article could be found in the Expression domain that applied to HR, and therefore we do not discuss it here. Additionally, the only Completion technique we could find a record for in the HRM literature was the sentence completion method, and therefore we focus only on this technique. In presenting these reviews and suggestions, we hope to encourage a line of research on projective measurement that embraces its purported advantages for HRM (e.g., resistance to faking, prediction of important outcomes), but avoids the pitfalls experienced by the pioneers of projective measurement in HRM (e.g., negative applicant reactions, legal action, philosophical deadlock).

2. Association techniques

Association techniques ask the participant to respond with whatever comes to mind after being presented with a stimulus, which is usually a picture or a word. This method is purported to “minimize ideation and emphasizes immediacy” (Lindzey, 1959, p. 163). In other words, the idea is to capture the subject’s immediate, unfiltered response. Building from Freud’s free association techniques, these responses are thought to be a window into unconscious motives, maladies, and conflicts. Here, we discuss two of the most popular association techniques that have been used in employment settings: the Rorschach and Word Association Tests.

2.1. Rorschach

Developed in 1921 by Swiss Psychiatrist, Hermann Rorschach, the Rorschach Ink Blot Test has remained one of the most well known (and controversial) projective tests ever used (Lilienfield et al., 2000). In the Rorschach, respondents are presented with a series of 10 inkblots (5 black and white and 5 with color) on separate cards and asked, “What might this be?” Frank (1948) proposed that these responses, once expertly scored, are indicative of internal emotional experiences that might normally be suppressed by the respondent. However, Exner (1989) rejected this “projective hypothesis” and argued that responses reflect the way in which the respondent conceptualizes the world. This is in line with Cattell (1951), who suggested that these techniques really measure respondents’ misperceptions, as opposed to their projections.

Table 1
Types of projective measures, examples of tests developed for clinical and organizational settings, and examples of past HRM uses.

<table>
<thead>
<tr>
<th>Type</th>
<th>Tests developed specifically for clinical settings</th>
<th>Tests developed specifically for organizational settings</th>
<th>Examples of past uses in HR-relevant areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association</td>
<td>• Cornell Word From (CWF; Mittelmann &amp; Brodman, 1946)</td>
<td>• Structured-Objective Rorschach Test (SORT; Stone, 1958)</td>
<td>• Selection of managers (Rorschach)</td>
</tr>
<tr>
<td></td>
<td>• Rorschach Ink Blot Test (Rorschach, 1912)</td>
<td>• Perceptanalytic Executive Scale (PES; Piotrowski &amp; Rock, 1963)</td>
<td>• Promotion of managers (Rorschach)</td>
</tr>
<tr>
<td></td>
<td>• Rorschach Group and Multiple Choice Test (Harrower &amp; Steiner, 1945)</td>
<td>• None</td>
<td>• Predicting supervisory performance (SORT)</td>
</tr>
<tr>
<td></td>
<td>• Draw-a-Person Test (Machover, 1949)</td>
<td></td>
<td>• Predicting managerial success (PES)</td>
</tr>
<tr>
<td></td>
<td>• House-Tree-Person Test (Buck, 1948)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Mirror Drawing Test (Brower, 1948)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Thematic Apperception Test (Morgan &amp; Murray, 1935)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td>• Washington University Sentence Completion Test (Loevinger, 1976)</td>
<td>• Office of Strategic Services Sentence Completion Test (OSS SCT; Murray &amp; Mackinnon, 1946)</td>
<td>• Assessment of implicit motives (TAT)</td>
</tr>
<tr>
<td></td>
<td>• Rozenzweig Picture Frustration Study (Rosenzweig, Fleming, &amp; Clarke, 1947)</td>
<td>• Personnel Reaction Blank (PRB; Gough, 1971)</td>
<td>• Assessment of performance (TAT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Miner Sentence Completion Test (MSCT; Miner, 1964)</td>
<td>• Prediction of occupational accidents (TAT)</td>
</tr>
<tr>
<td>Completion</td>
<td>• None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrangement/Selection</td>
<td>• Szondi Test (Szondi, 1947)</td>
<td>• Tomkin–Horn Picture Arrangement Test (PAT; Tomkins &amp; Miner, 1957)</td>
<td>• Holistic Selection of high-level military personnel (OSS SCT)</td>
</tr>
<tr>
<td></td>
<td>• Luscher Color Test (Luscher &amp; Scott, 1969)</td>
<td></td>
<td>• Mechanical Selection of lower-level employees (PRB)</td>
</tr>
<tr>
<td></td>
<td>• Projective Puppet Play (Wolffman, 1960)</td>
<td></td>
<td>• Predicting counterproductive work behavior (PRB)</td>
</tr>
<tr>
<td>Expression</td>
<td>• Handwriting Analysis (Beyerstein &amp; Beyerstein, 1992)</td>
<td>• None</td>
<td>• Selection of managers (MSCT)</td>
</tr>
<tr>
<td></td>
<td>• None</td>
<td></td>
<td>• Promotion of managers (MSCT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Attitude change in managers (MSCT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Executive personality and motivation assessment (PAT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Selection of salespersons (PAT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• None</td>
</tr>
</tbody>
</table>


As with most other assessment techniques of its time, the Rorschach test came under increased scrutiny and criticism in 1950s due to a lack of scoring standardization procedures, a lack of normed data, low interrater reliability, and weak validity evidence (Dawes, 1994). In fact, Dawes noted, “...in 1959, many of the world’s most eminent psychologists were lined up against the use of the Rorschach” (pg. 151). In response to these criticisms, Exner (1974) developed the Comprehensive System (CS) for scoring the Rorschach, which standardized the administration and scoring procedures and also offered norms for both adults and children. In this system, responses are scored based on factors related to content (e.g., people or sexual content), location (e.g., focus on the whole inkblot vs. a specific area), determinants (e.g., observing something involving color, dimensionality, or movement), and other characteristics (e.g., the organization of the response, illogical and uncommon responses). These scores are then combined in a structural summary, which provides standing on broad dimensions such as depression, egocentricity, and ability to cope with stressors.

With regard to psychometric properties of the Rorschach’s CS, there remains a great deal of controversy. For example, interrater reliability evidence has been shown to be highly variable across the various CS scores. In fact, one study found that intraclass correlation coefficients (ICCs) of two graduate student raters ranged from .2 to 1.0 with a median of around .80, depending on the CS dimension (Acklin, McDowell, Verschell, & Chan, 2000). Although this points to problematic reliability evidence for some dimensions, a more comprehensive study by Meyer et al. (2002) found ICCs ranging from .82 to .95 depending on the type of rater, with a mean pooled ICC of .91. The temporal stability of the Rorschach appears to be less contentious. In a meta-analysis of CS studies that reported test–retest reliability, Grennerod (2003) reported correlations of .84 over three weeks and .5 over five weeks. However, more research is needed in this area as it was found that only about 40% of Rorschach studies report test–retest reliability (Garb, Wood, Nezworski, Grove, & Stejskal, 2001). Validity evidence for the CS, mostly conducted in the clinical realm, has also been mixed and highly dependent on the dimensions and criteria chosen. Across dimensions, meta-analyses tend to report mean validity coefficients of about .30 (Lilienfeld et al., 2000). Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neuleib (1999) found that the type of criterion measure (self-report vs. objective criterion) was a significant moderator. That is, the MMPI was more strongly related to self-report outcomes and the Rorschach was more strongly related to objective outcomes. The authors reason that the larger correlation between the Rorschach and objective outcomes may be due to the inability of respondents to respond in socially desirable ways on the Rorschach.

The use of the Rorschach in HRM settings has been much more sparse, and most research done to predict work outcomes was conducted during the 1940s, 50s, and 60s. For example, Dulsky and Krout (1950) used the Rorschach to successfully predict promotion potential in a sample of factory supervisors. Corresponding with the drastic reduction in research attention, the Rorschach’s use in applied settings as a selection tool has also dramatically decreased. In Spielberger, 1979 posited that the Rorschach was one of the most frequently used selection tests in police departments. However, research has shown that it has been largely abandoned over the years with a more recent survey finding that only 5.8% of police departments used the test in 2003 (Cochrane, Tett, & Vandecreek, 2003).

Piotrowski and Rock (1983) developed a complex method for coding the Rorschach, called the Perceptanalytic Executive Scale (PES), which was used to identify talent for managerial positions. This scale consisted of 32 discrete test reaction scores combined to form an overall score of management potential. In their study of 110 executives, independently categorized as either successful or unsuccessful, analyses showed that scores on the PES accurately predicted managerial success. Given the purported initial success of this scale, it is surprising that the PES was never widely implemented in managerial selection systems. This could be due to a lack of replication, the development of the CS, or the difficulty of scoring (Del Giudice, 2010).

Although the Rorschach is purported to provide a rich and in-depth understanding of employee personality structure, it is also cautioned that the Rorschach may be too expensive and time-consuming to administer, especially to non-managerial employees (Brower & Weider, 1950; Reiger, 1949). We see this as is a major drawback of this method and a likely roadblock to its further development for employment settings. To address these limitations, a group administration and multiple-choice version of the test were also developed (Harrower & Steiner, 1945). The validity evidence for the group administration method has been mixed (see Jensen & Rotter, 1945; Piotrowski, Candee, Balinsky, Holtzberg, & Van Arnold, 1944).

A variant of the Multiple Choice Rorschach was developed, called the Structured-Objective Rorschach Test (SORT; Stone, 1958). This measure specifically aims to identify temperaments that are related to organizational functioning. Among others, these include theoretical, pedantic, activity potential, and human relationships. Hicks and Stone (1962) found that the S-O Rorschach was the strongest predictor of several performance outcomes in a battery of self-report measures given to 76 supervisors. For example, the human relationships dimension was significantly correlated with peer ratings (though not supervisor ratings) of performance (.39), promotability (.29), and versatility (.25).

It should be noted that these studies used different dimensions of the Rorschach, thus limiting their comparability with regard to validity evidence. Further, the limited research examining the psychometric properties and validity of more practical Rorschach formats (e.g., SORT) for use in organizational settings has produced mixed results. Therefore, future research may be needed to determine the validity of the Rorschach in organizational settings. Del Giudice (2010) has taken a first step by outlining the ten Rorschach dimensions that have been shown to be the most strongly related to leader performance (e.g., cognitive capacity, self-improvement, autonomy). As we note later, there may be other issues that preclude the utility of such research.

2.2. Word association tests

The word association technique relies on a stimulus word to which respondents are asked to respond with the word(s) that first comes to mind. This technique has been in use for over a century. Sir Francis Galton (1879) is frequently credited as the first
to use word association to study his own reactions to a set of 75 vocabulary words. James McKeen Cattell, inspired by the work of Galton (see Gillham, 2001), was also an early pioneer of this method of measurement. Cattell and his colleagues conducted continuous collections of associative responses to words in his laboratory over many years.

Although Galton and Cattell’s work on word association was important, their studies dealt mostly with reaction time and the cognitive aspect of associations. It would be the psychodynamic perspective of Jung, 1910) that would be most influential in shaping the technique as a projective measure of internal, non-cognitive variables (see Klopf, 1973). In 1909, Jung presented a talk at Clark University on the topic. In this talk, Jung presented 100 words meant to “strike easily almost all complexes or practical occurrence,” presenting differences between response times across words for normal and hysterical persons. Further, he discussed scoring in a statistical fashion that might surprise some modern applied psychologists.

Most importantly, Jung was the first to show that associative responses yield not only information concerning cognitive process, but also emotional matter. Jung and Eder (1919) published an entire book on the topic of association, along with studies concerning response times and attempts to distinguish “normals” from those with psychological complexes. Following this work, word association tests saw an initial surge in interest, but experienced a large decrease in popularity around the 1950s (see Vernon, 1953).

Projective word association techniques have seen some interesting applications to HRM. Early uses can be found in brief descriptions of military applications, such as selection of fitness for flight training (Kellum, 1948), the selection of British Army Officers (Ferguson, 1947), the development of the in-house AirCorps Word Association Tests (US War Department, 1941), as well as the Kent–Rosanoff test in American Army medical installations. However, these brief accounts tend to either provide little information (perhaps due to security concerns) or discuss word association tests as ancillary to other measures (e.g., TAT and Rorschach; see US War Department, 1941).

At least one author was hopeful for the use of word association tests in more general HRM applications. Brower and Weider (1950) touted the Cornell Word Form (CWF; Mittelmann & Brodman, 1946) as a high-potential measurement technique, with particular attention to its presumed resistance to faking. In the CWF, the respondent is asked to provide two responses to each stimulus word (e.g., sleep, mother). Weider (1951) showed significantly different CWF scores for employees reliably rated “good” and “poor” in their performance. However, the remainder of his studies appears to have been more medical and psychiatric instead of HRM-oriented (e.g., Weider, Keeye, Bela, & Wolff, 1946).

Gough (1976) provided an interesting application of word association tests to the measurement of creativity in a sample of research scientists and engineers (most with PhD or Master’s degrees). In this article, the Kent–Rosanoff test (Kent & Rosanoff, 1910a,b) was compared to Gough’s own word association test. Whereas the Kent–Rosanoff is a collection of 100 seemingly random words which was designed for diagnostic use in clinical settings, Gough used a measure composed of 100 scientific terms as stimuli (e.g., “protons”). Participants responded to both measures such that they provided the first word to come to mind. For both measures, different samples were used for norms by tabulating the frequency of each of the free responses given by participants, categorizing word usage rates. Gough hypothesized and confirmed that being in the 1% to 9.9% range was predictive of peer and supervisory ratings of scientists’ creativity on the scientific word association test, but to a lesser extent for the Kent–Rosanoff. The author concluded that moderately uncommon responses were indicative of creativity, whereas very uncommon responses (<1%) actually indicated the opposite.

Surprisingly the Gough study represented the only in-depth study of word association tests applied to an HRM-related area that we were able to locate. Molish (1972) speculated the decline was at least partially due to the researchers’ shying away from the psychodynamic perspective and issues involved with response sets, norming, and language (Rabin, 1968). Therefore, although Gough’s (1976) study showed that usage rates were related to creativity in scientists, little is known about how such techniques would fare in more general HRM applications and settings.

3. Construction techniques

Construction techniques require the subject to create something, such as a story or drawing (Lindzey, 1959). This response requires more cognitive processing than word associations. Because of the extremely ambiguous nature of the stimulus, subjects are thought to project inner qualities in the construction of the response. We discuss the most common (by far) construction technique used in employment settings: the Thematic Apperception Test (TAT).

3.1. Thematic Apperception Test

The TAT was first developed by Henry Murray at the Harvard Psychological Clinic in the early 1930s. The TAT consists of a set of ambiguous pictures about which subjects are asked to create a story (Morgan & Murray, 1935). These stories are assumed to reveal important information about the motivations and personality of respondents. Murray’s (1938) seminal book, Explorations in Personality, included some experimentation using the TAT. However, the book was mostly an exploration in how to best use the method, and ways to cull some meaning from it through subjective interpretations on the part of the clinician.

Later researchers would seek to gather validity evidence using more empirical approaches. Specifically, David McClelland and his colleagues – most notably John Atkinson – developed a program of research based on theories of basic human motives. For example, to assess validity they experimentally induced the drive of hunger in participants, and then compared self-reports of hunger with a TAT that was content-coded for hunger themes (Atkinson & McClelland, 1948). They found that self-report ratings for hunger were dependent on both time of day (a social/external cue) and the number of hours elapsed since participants had eaten, whereas the TAT stories told by participants were dependent only on the length of food deprivation.
These authors also worked to refine the scoring procedures of the TAT to make them more reliable and valid (Atkinson & McClelland, 1948; McClelland, Atkinson, Clark, & Lowell, 1953). Whereas Murray believed that the scores from each “expert” rater of the TAT should be pooled to form a composite score, Atkinson and McClelland argued that this compounded the error implicit in each person’s subjective rating, and created a scoring that measured motive arousal for three primary needs: n Achievement, n Affiliation, and n Power.

Meta-analytic correlations between TAT-based, unconscious motives and self-report, conscious motives are null (McClelland, Koestner, & Weinberger, 1989; Spangler, 1992). McClelland argues that this reflects the strength of the TAT, as it indicates the two are tapping different aspects of the same construct (i.e., implicit vs. explicit motives; McClelland, 1999). McClelland also argued that early accounts of poor test–retest reliability that resulted in the popular view of the TAT as unreliable, may have been due to the instructions given to “be creative” when developing stories. He argued that this caused participants to work harder on the second measurement occasion to diverge from their original story for the sake of creativity. When participants are given specific instructions to not worry about the level of similarity or difference between the stories, test–retest reliabilities are significantly higher compared to instructions to make the stories different, though still low by self-report standards (Winter & Stewart, 1977).

In the first published use of the TAT applied to personnel testing, Murray and Stein (1943) discuss the conduct of a selection battery to select United States Army officers for combat duty by measuring their potential for leadership. The measure, the Rapid Projection Test (RPT) was described as an “offspring” (p. 389) of the TAT meant to be administered in both small and large groups. When groups were small, examinees viewed twelve images containing one person who was matched to their sex. Examinees were then prompted by two questions regarding each image, and were told this was a test to see how well they could “size up people.” The first, “What is he thinking?” was administered for every picture. The second could be, “What has happened?” “What is he feeling?” “What will he do?” or “How will it turn out?” For each of these two questions, respondents could choose from six answers to these questions, and a sum-score was used to measure twelve traits (one per picture).

Briggs (1954) presented a modification of the TAT for use in group assessment, consisting of ten cards made particularly to select enlisted Navy for admission to the U.S. Naval Submarine School at New London. This report was followed by another Navy report that explicated the inter-rater agreement (between .81 and .84 Pearson r between raters, and around 65% agreement), but provided no correlations or findings regarding external variables (Eron, Sultan, & Auld, 1955).

By 1966, a retrospective on the use of projective measures in personnel psychology was presented by Kinslinger in Psychological Review. He notes some interesting studies not mentioned above, including findings that indicated significant differences in TAT scores reflecting Need for Achievement for salesmen categorized as high and low–performing based on supervisor ratings (Botha & Koper, 1963). More notably, Kinslinger (1966) discussed a rare type of criterion-related validity study showing lower accident liability for a group of South African bus drivers selected using the TAT and a group not selected by the TAT (Shaw, 1965).

By the 1970s it appears work in HRM-related domains had slowed considerably. We were only able to find a handful of studies from after this point. Durand and Shea (1974) showed that Need for Achievement was positively related to the level of business activity 18 months following assessment in a sample of persons operating small businesses. In an article reflective of the more recent move of industrial–organizational psychologists to person-organization fit, Johnson (1974) found that those high in Task Orientation as measured by the TAT had a more positive self-reported perception of their relationship to the organization. Morris and Fargher (1974) showed that in a sample of 60 business owners, high n achievement TAT scores were significantly related to the growth rate of their respective businesses.

Mohan and Brar (1986) showed in a study of 93 Swiss trainees and 101 skilled industrial workers that there were no significant differences between high and low work efficiency groups in regards to their TAT scores for the three needs. Spangler (1992) found that a TAT n Achievement measure strongly predicted career success in the presence of intrinsic or achievement incentives, whereas questionnaire measures were positively related to career success in the presence of extrinsic or socially-influenced incentives. Overall, TAT relationships were higher. On the other hand, a meta-analysis by Collins, Hanges, and Locke (2004) found that both TAT and questionnaire measures predicted entrepreneurial behavior with no significant difference in validity coefficients between methods. Perhaps the most recent study relevant to the concerns of HRM scientists was Dreyfus (2008), who showed the TAT had no predictive utility in explaining the effectiveness of research and development managers.

In spite of the lack of fully consistent evidence for differences in criterion-related validity, it seems that there is only some support for McClelland’s assertion that the two measurement methods tap different aspects of the same construct domain. For example Spangler’s (1992) findings of differential validity provides support for the idea that implicit measures predict people doing things, especially if they are non-declarative, whereas more overt measures predict people’s conscious descriptions and are likely to relate to more declarative behavior. However, the findings of Collins et al. (2004) have found no differences in the two methods’ criterion-related validities relating scores to overt behavior. Adding to this complexity, some studies discussed above found no evidence for the TAT predicting important HRM-relevant outcomes, whereas others have. Thus, although it appears that the two measures may tap different aspects of the same construct, more research would be needed to disentangle this wide variety of findings. However, as we will note later, there may be limited utility in these and other more clinically oriented measures for HRM practitioners.

### 4. Completion techniques: sentence completion tests

SCTs are likely the most commonly used (in any profession) of the projective measures (Holaday, Smith, & Sherry, 2000). For SCTs, respondents provide the rest of a partially completed sentence. In spite of their genesis being related to the word association
measures described here, they have avoided some of the dogma surrounding the psychodynamic perspective after breaking with it. In fact, in the same review that noted the aforementioned perception of word association, the author notes that researchers believe SCT approaches to be among, if not the most, valid of projective techniques (Molish, 1972). Further, SCTs represent a unique instance of projective techniques, in that measures designed specifically for personnel and HRM management purposes are abundant relative to other projectives.

Similar to word association methods, the SCT approach began as a way to measure cognitive variables. Herman von Ebbinghaus was famous in his early career for work on human memory; however, in later years he expanded his repertoire and was involved in research involving mental tests that would later inform the Stanford–Binet intelligence tests. Part of this research involved designing an SCT (Ebbinghaus, 1897), which was later adapted for use in the English language (Kelley, 1917; Traube, 1916; Ebbinghaus wrote in German). However, it would not be until 1930 that the SCTs were used as projective measures (Symonds, 1947).

The first instance of the SCT method as a projective was Tendler’s (1930) research designed for gaining insight into emotions. The article contains no references and begins by discussing the disadvantages of free association and self-report forms. Tendler’s approach to SCTs was psychodynamic in perspective in the same way as word association techniques, which he notes as a major advantage of the approach for both clinical and personnel work. Despite Tendler’s exposition of the advantages of this approach and showing some validity evidence using self-report personality inventories, Tendler’s appeal has been mostly ignored (see Symonds, 1947).

Perhaps the first large-scale application of SCTs to HRM-related areas was in the Office of Strategic Services (OSS) Assessment Program by Murray and MacKinnon (1946), which was designed to select candidates into the OSS. Although Murray and MacKinnon only mention the SCT in passing, the method was the focus of two works by Morris Stein, a clinical psychologist at the Veterans Administration (Stein, 1947, 1949) whom suggested it could be used instead of Rorschach and TAT techniques, especially for unimaginative persons. However, as discussed by Highhouse (2002), little statistical evidence was used in the OSS program, and information from psychologists’ general reports was integrated holistically to make decisions. Although some general research on SCTs was stimulated by the OSS examinations (e.g., Rohde, 1946, 1947), most of this research through the 1960s would concern applications to clinical diagnostic issues (e.g., Aronoff, 1967; Forer, 1950; Rotter, Rafferty, & Schachtitz, 1949; Sacks & Levy, 1950). However, two specific measures, the Personnel Reaction Blank and the Miner Sentence Completion Scale, would come to see attention in the HRM area.

4.1. Personnel Reaction Blank

The first focused application of an SCT in HRM-relevant research can be credited to social/personality psychologist Harrison G. Gough. Gough was a professor of psychology at the University of California, Berkeley from 1949 to 1986, and is most famous for developing the California Psychological Inventory (Gough, 1956) and co-founding Consulting Psychologists Press (now CPP, Inc.). However, he was also the creator of the Personnel Reaction Blank (PRB), a measure created on the basis of social psychological theory as opposed to the psychodynamic measures of the past. The PRB has been noted as the first personality-based integrity test, and has been noted to have an exceptional combination of advantages in psychometric and economic terms (Viswesvaran & Ones, 1997).

Gough developed his theory of psychopathy in 1952 (Gough & Peterson, 1952) and applied this theory to stimulus response through the PRB. Under this view, antisocial individuals would have difficulty taking the perspective of others. He posited this inability was indicative of a predilection to have a lack of self-control, and therein to exhibit a pattern of antisocial behavior. The PRB asks respondents to complete vague statements by indicating whether they are True or False. The item content is vague in that it is not obvious that the items tap the likelihood of counterproductive work behaviors.

The PRB was first developed by Gough as a non-commercial research tool in 1954 to measure dependability, conscientiousness, and social conformity (Gough, 1971: p. 669). The first application of the (then) 94-item measure noted by Gough was an Italian translation and standardization of the form (Gough, 1962), vaguely described as an application to personnel selection (Sciola, 1962). Loudermilk (1966) appears to be the first to use the test. He studied the relation between batteries of physical and psychological assessments (e.g., aptitude, personality) administered to workers in paper and lumber mills. Among the 13 resulting test scores, the PRB showed the largest correlation with supervisors’ ratings of work efficiency, \( r = .33 \), rivaled only by measures of Verbal and Numerical Aptitude. Gough (1965) also authored papers in the Italian journal, Bollettino di Psicologo Applicata, using the PRB. One study of Italian clerks showed a correlation of .28 between the PRB and performance ratings, and another (Gough & Freddi, 1967) showed a correlation of .38 with performance ratings of 79 steel workers.

In Loudermilk’s (1966) publication, the measure is described only as a “short personality test developed by Gough (1954)… to assess the ‘dependability-conscientiousness’ personality factor” (p. 303). However, by its next appearance in the literature in 1971, the PRB had gone through iterations wherein it was a 92-item and finally a 70-item measure of all three constructs originally mentioned by Gough in 1954 (see above) that was made commercially available (Gough, 1972). In the first published use of the English form by Gough (1971), he compared a large sample of delinquent (i.e. in jail or juvenile detention, \( N = 3,034 \)) and non-delinquent (\( N = 619 \)) Americans that suggested the PRB showed high discrimination between these subsamples for both males and females.

Note the OSS would later become the United States Central Intelligence Agency.
After 1971, very few applications of the PRB appear in the literature until the 1990s. To date, this research has shown mostly positive results. From our review of the literature, it is apparent that this revival of the PRB was due to the influential review of the honesty/integrity test problem by Sackett, Burris, and Callahan (1989). In fact, the rate of PRB publications from 1990 to 1999 more than doubled in comparison to the number studies published prior to that. However, around 1996, the popularity of the PRB appears to have experienced a major decline. We believe it is likely that evidence of negative applicant reactions (Whitney, Diaz, Mineghino, & Powers, 1999) and faking (Alliger, Lilienfeld, & Mitchell, 1996) may be the reason for this reversal.

In 1989, Frost and Rafilson compared what they refer to as overt (i.e., typical self-report with more transparent item content) and covert personality-based tests. The authors purported that an overt test (the Honesty scale of the Personnel Selection Inventory; London House, Inc., 1980) was better at predicting theft, counterproductive work behavior, and drug abuse than the covert PRB. However, a major limitation was that outcome measures were self-report. Therefore, in our opinion it is likely that the outcome is actually related to brazen counterproductive behaviors, and could actually indicate support for the validity of the PRB for predicting more covert (as opposed to overt) behaviors. This is consistent with findings involving other projective measures (see our discussion of the TAT). In fact, using factor analysis, Woolley and Hakstian (1992) showed that the PRB hangs together with a group of other covert personality-based integrity measures that differ distinctly from the overt measures of similar constructs used for similar purposes (i.e., employment testing). This makes the finding of Frost and Rafilson (1989) somewhat less surprising.

Using a more high-fidelity outcome, Kobbs and Arvey (1993) showed that subsamples of nurses who did and did not have official action taken on their licenses were well-discriminated by scores on the PRB, even when controlling for factors such as education, sex, and household income. Contrary to Frost and Rafilson (1989), another investigation by Woolley and Hakstian (1992) showed the PRB was related to admissions of counter-productivity. Further, the results of Connelly, Lilienfeld, and Schmeelk (2006) showed that the PRB was unrelated to mental ability, moral reasoning, and ego development measures, and was negatively related to measures of several psychopathetic personality traits (e.g., Machiavellianism). The authors concluded that this latter finding suggests integrity tests are more likely to be associated with actual moral behavior than with moral reasoning, a positive finding for users of integrity tests.

Although much of the early research on the PRB was positive, some more recent investigations appear to take a more negative tone. Whitney et al. (1999) showed that applicant reactions toward personality-based integrity measures (including the PRB) were more negative than self-report forms. Further, recent research has shown that respondents can successfully fake the test (Alliger et al., 1996; Byle & Holtgraves, 2008), lending skepticism to one of the more prominently purported advantages of the PRB.

In its most current form, the personality aspect of the PRB contains 62 scored items that sum to form the Person Reliability Index (PRI), which is made up of four sub-scales: a) Sense of Well-Being; b) Prosocial Background; c) Compliance with Social Norms; and d) Conventional Preferences. Most recently, this updated measure has been used to argue that a focus on externalizing, a trait related to impulsivity, antisocial behavior, and risky behavior including substance abuse, should be considered by integrity test researchers (Blonigen et al., 2011).

In sum, the PRB has shown a good deal of evidence that supports its use in HRM-related areas for predicting important outcomes such as performance, and has shown an ability to distinguish between persons likely and unlikely to engage in and endorse undesirable behaviors. However, more recent research has cast doubt on its utility. Particularly, evidence of the potential for faking and concerns over applicant reactions appears to have driven researchers away from its consideration in HRM applications. Alternatively, more recent versions of the PRB and the work of Blonigen et al. (2011) suggest that the PRB may have the potential to make a comeback in HRM.

4.2. The Miner Sentence Completion Test

Whereas the PRB was designed to assess lower-level employees, another SCT popular in HR-related arenas is the Miner SCT (MSCT; Miner, 1960), which was developed for managerial assessments involving selection and promotion (Berman & Miner, 1985; Carson & Gilliard, 1993; Ryan & Sackett, 1987). The Miner SCT (MSCT) was developed by John B. Miner, then a professor at the University of Oregon. Miner is well-known as a leading theorist on management, and has been highly involved in research on the MSCT for decades.

The MSCT asks respondents to complete a series of sentences that are then scored by judges as “Positive,” “Negative,” or “Neutral.” Example item stems are “Policemen...” and “My family doctor...” At formulation, the MSCT was described as a measure of attitudes toward supervisory jobs, but Miner began to believe it more generally tapped a “desire to compete for and exercise power over other people” (1961, p. 18), later conceptualized as a motivation to manage (Miner, 1964). By 1965, Miner had further refined the framing of the MSCT within the role-motivation theory of managerial effectiveness (e.g., see Berman & Miner, 1985), which he frequently noted as being similar to McClelland’s theories of achievement motivation (see McClelland, 1961; McClelland & Winter, 1969; Miner, Smith, & Bracker, 1994). This theory posits that managerial positions in different organizations have several roles in common. Correspondingly, the MSCT measures the various roles posited by role-motivation management theory (e.g., attitude toward authority figures and motivation for competition in games). For a complete review of the theory see Miner (1993).

Miner’s (1963) first study using the desire-based conceptualization of the MSCT (Miner, 1968) showed that, as predicted, a group of research managers were lower in the desires discussed above than sales managers. A series of studies by Miner and his colleagues would follow, further aligning the MSCT with role-motivation theory and pointing out some of the limitations of the
theory and method. In 1968, Miner claimed validity evidence for the MSCT by showing those with explicitly stated managerial goals showed significantly higher motivation to manage than those without such goals. However, in 1971, Miner presented a series of three studies showing that the MSCT did not have any predictive utility in predicting success of consultants, whereas the Tomkins–Horn Picture Arrangement test (discussed below) did. His conclusion was that the MSCT was likely more valuable in larger organizations (e.g., Miner, 1963), and not in small, professional consulting firms (e.g., Miner, 1971).

Both Miner (1965) and Bowin (1973) showed that students taking a course on managing ineffective performance (based on role motivation theory) positively changed their attitudes (measured by the MSCT) toward role prescriptions assigned at the beginning of the course, suggesting support for the measure’s validity (Miner, 1993). Showing another boundary condition of role management theory, Miner, Rizzo, Harlow, and Hill (1974) used large samples of students in simulated organizational situations to show that MSCT scores were related to promotion decisions in highly structured, bureaucratic organizational simulations, but not in low-structure conditions. This supported Miner’s (1971) conclusions regarding his findings in the sample of consultants discussed above.

Although much of Miner’s findings for the MSCT were positive, some researchers remained unconvinced of the approach. Brief, Aldag, and Chacko (1977), taking issue mostly with the subjectivity of ratings provided to responses, presented an investigation of the MSCT in relation to two self-report measures and concluded that there was a lack of evidence of convergent and discriminant validity with these measures. Miner (1978) responded to these allegations. First, Miner showed that more experienced researchers showed higher agreement than those in the Brief et al. (1977) study. In responding to construct validity criticisms, Miner asked “...whether a failure to find significant relationships between the MSC[T] and the PVQ [Personal Values Questionnaire] has anything to do with the construct validity of either instrument. One is a measure of motives, the other of values... Because two variables are consistently correlated with a third [managerial success] is not reason to expect that they should necessarily correlate with each other” (Miner, 1978, p. 288).

One interesting feature of the exchange between Miner and Brief et al. involved the possibility of using a multiple-choice version of the MSCT. In 1977, a multiple-choice version of the MSCT was introduced (see Miner, 1977), and Brief et al. (1977) recommended it replace the original based on their findings. However, Miner (1978) noted that the multiple-choice version, although predictive of managerial success, had shown high score inflation relative to the free-response version. As shown above, several replications of Miner’s (1965) original result had confirmed his findings in various conditions with the original MSCT, and two studies would continue this trend for the MSCT with minimal race and sex differences (Bartol, Anderson, & Schneier, 1981; Miner & Crane, 1981). However, things suddenly changed around this time.

In 1981, a commercial version of the MSCT was published that included a component with a multiple-choice format. The original scale remained unchanged and was designated Form H (for hierarchy, referring to the hierarchical structure of the MSCT), with the new form being termed Form P (for use in professional systems). In the next few years, some controversy in scoring the MSCT saw light in the published literature. This controversy was mostly between Miner and Bartol, whose investigations had shown some major inconsistencies, most importantly, a disagreement in the decline in managerial motivation found by Miner and Smith (1982). Although we do not review this debate in detail, the basic question was the effect of raters, with Bartol, Schneier, and Anderson (1985) arguing that Miner’s raters, although more reliable, were possibly biased in their ratings (giving higher ratings on average than Miner himself). An attempt at discrediting a similar argument was lodged by Miner, Smith, and Ebrahimi (1985) against Bartol, Anderson, and Schneier (1980).

After this controversy, research using the original, free–response form of the MSCT slowed considerably, with Bartol turning to use of the forced-choice version (e.g., Bartol & Martin, 1987), and then others (Erberhardt, Yap, & Basuray, 1988). In spite of some early use of the forced-choice version form, the form appears to have never really “taken off” in that only these two studies were found using this form.

In the years following 1990, mostly meta-analytic results for the MSCT were published, with some exceptions that were still establishing the basic hypotheses addressed by Miner in 1965 with variations on questions related to race (e.g., Stevens & Brenner, 1990; Tomkiewicz, Brenner, & Esinhart, 1991). Carson and Gilliard (1993) showed high and stable (i.e., low variability) correlations of just over .30 with performance ratings, and variables suggesting “opting in” to management professions (i.e., managerial status, career plan). Eagly, Karau, Miner, and Johnson (1994) showed significant, but small, gender differences on the original projective form, using 51 datasets covering a 30-year period. A meta-analysis involving the relation between achievement motivation and entrepreneurial activity showed that the MSCT had higher correlations with activity than the TAT and for both projectives to be “on par” with one another in prediction, but were uncorrelated with one another. As the authors note, this may be suggestive that certain measures represent different aspects of the construct (Collins et al., 2004), a position we will consider in detail in the Discussion section.

Although the MSCT has shown mixed results, the majority of studies showed favorable results in a wide variety of settings using highly variable criteria. Further, we believe many of the less positive findings for the projective form were likely misleading, due to researchers’ consideration of correlations with self-report measures as a necessary condition for validity, a similar complaint lodged against the TAT. As will be addressed in our discussion, we believe this represents an unfortunate state of affairs in research on projectives applied to HRM. Overall, given the mostly positive findings for the MSCT and the importance of motivational constructs to the management literature, we take the position that the slowing of research involving this measure was likely unwarranted.

5. Arrangement techniques: The Tomkin–Horn Picture Arrangement Test

Arrangement techniques require respondents to order a series of pictures or objects based on some criterion, such as correctness, attractiveness, or relevance (Lindzev, 1959). It contains elements of association because the various options are
compared, as well as elements of construction because the ordering often tells a story. Here we focus our discussion on the one arrangement technique developed specifically for HRM purposes, the Tomkin-Horn Picture Arrangement Test (PAT; Tomkins & Miner, 1957).

The PAT requires the subject to arrange a series of three ambiguous pictures in a way that makes the most sense. They are also asked to write a sentence for each of the three pictures to explain what they believe is happening in the story. The full version contains 25 such tasks, though versions with as few as 3 have been used. All of the pictures involve people engaging in daily activities, with over half having something to do with work. There is no clear, expected order in which the pictures should be arranged, allowing the respondent freedom to create a story through their chosen arrangement. Some of the characteristics purported to be measurable with this technique include aggression toward supervisors, persistence and consistency of work habits, and reaction to criticism (Brower & Weider, 1950).

In Miner and Culver (1955) analyzed data from 44 top-level executives and 31 professors and found that the executives’ arrangements were more likely to suggest fears of illness and failure, as well as feelings of helplessness. One study investigated the predictive validity of the PAT as well as several other personality/cognitive ability measures in a sample of salesmen. Using several indices of objective performance, the author found that the PAT correlated in the high .5 range, considerably higher than the Wechsler Adult Intelligence Scale and confirmed this relationship through cross-validation (Miner, 1962), making this a particularly strong study. A study in 1973 found that management consultants scored higher than managers on the PAT-rated dimensions of work motivation, nonconformity, and emotional instability (Miner, 1973).

Though there appeared to be some promise regarding the validity and usefulness of the PAT, it is no longer used in organizations in its original form (Highhouse, 2002), and we could find no more published applications than the handful of studies summarized above. Further, we could find no clear reason for the decline in its use, though it may have been due to Miner’s shift in focus to the MSCT (see above), which was more practical from an administration perspective.

6. Discussion

As can be seen, a good deal of research has been conducted using projective techniques in application to HRM problems. Results have been mixed for many measures, whereas others have had more positive outcomes. From our literature review, we view projective techniques as having serious promise in terms of utility for application, as well as construct validity that is important for both research and application. However, a common theme among all the projective techniques is a general decrease in HRM applications. This seems unfortunate given the susceptibility of self-report measures to faking and more recent advances in projectives that are discussed elsewhere in this issue (e.g., conditional reasoning; implicit associations).

We posit that this decrease has likely resulted for two major reasons. First, practical and legal complexities arise when considering the use of projectives. It is our position that most of these concerns could be alleviated if HRM were to work toward building its own literature concerning both the development and application of projective measures. Secondly, it would appear that researchers at some point in the study of projectives in HRM contexts reached a sort of philosophical deadlock. By this, we mean that researchers of one ilk were not convinced by the methodology or logic of another, and vice versa. Below we discuss these issues in turn, and try to make suggestions where appropriate.

7. The need for an HRM-centric literature on projectives

First and foremost, we believe there is a pressing need for an HRM-centric literature on projectives. This position grew out of our observation that most research on projective measures in HRM applications originates from outside of the field. Although it is wise to draw upon general psychological theory for informing measure development, an unfortunate fact is that nearly all the measures discussed here were originally meant for, or developed within, theoretical traditions rooted in clinically oriented realms. As will be noted, this could possibly carry legal implications, which is not unrelated to the next point: HRM researchers have very different practical considerations than clinicians and research psychologists. Below, we detail some potential problems with adopting measurement methods developed for clinical psychology. We then note that an HR-centric literature could alleviate most, if not all of these problems.

7.1. Measures should be based on “normal” theories

What is clear from our review of projective testing is that nearly all were developed under frameworks designed for detecting psychopathology. Even many of the projective tests developed specifically for the purposes of HRM are still based on clinical conceptions of abnormality. For example, whereas a particular scoring system was developed for the Rorschach (i.e., the Perceptanalytic Executive Scale) to select managers, the Rorschach itself was developed to identify abnormal psychological functioning. In addition, the PRB was initially validated on the outcome of delinquency and was based on Gough’s theory of psychopathy (Gough & Peterson, 1952). To a certain degree, these are outcomes that are of interest to HRM professionals (e.g., stress, negative self-perceptions, and depression). However, it would be better to develop techniques using “normal” theories of human behavior and personality. For example, developing a less fakeable measure of conscientiousness would be of great benefit to the field. This is also in line with calls for more of a focus on positive psychology in the management sciences (Luthans, 2002; Zickar, 2010) in that it would focus on what is “right” with individuals as opposed to what is “wrong.”
Moving away from clinical-based measures is highly important from a legal standpoint. The Americans with Disabilities Act prohibits employment discrimination on the basis of mental disability. Thus, the use of anything considered a medical test of mental disability would be off-limits for organizations. The EEOC outlines factors that indicate whether a test is of a medical nature or not. These include, “whether a health care professional administers the test and analyzes the results, whether the test is meant to determine impairments or evaluate health, whether the test is invasive, whether it tests physiological characteristics or task performance, whether it takes place in a medical environment, and whether the administration of the test requires the use of medical equipment” (Gonzales-Frisbie, 2006, p. 191).

Based on these guidelines, it seems the most problematic features of projective testing in the workplace is that most tests were originally developed to measure some form of mental impairment. Indeed, recent case law ruled that a personality assessment can violate employment provisions of the ADA if the test was originally designed to detect mental illness. In *Karraker v. Rent-A-Center*, the use of the MMPI was ruled in violation of the ADA even though only normal personality was being used for selection purposes (see Gonzales-Frisbie, 2006). It was reasoned that even though it was not the intention of the defendant, low scores on the MMPI were more likely to indicate disability. Additionally, the invasiveness of clinical-based measures could violate the ADA. Affirming this issue, the ruling in *Soroka v. Dayton Hudson Corporation* (1991), the MMPI was found to violate California privacy laws. The issue of invasiveness is certainly applicable for even HRM-centric measures grounded in theories of psychopathy. For example, *Jones (1991)* found items from the Employment Inventory and the PRB were rated as more invasive than the more job-relevant questions in the Personnel Selection Inventory. Although the case is not directly relevant, the Supreme Court’s guidelines resulting from *Ricci v. DeStefano* (2009) should be considered for developing projectives that include job-relevant stimuli. Finally, we believe Gough’s (1976) use of science-related stems and responses in studying creativity in scientists can be considered an exemplary method of development and standardization of a job-relevant projective measure.

7.2. HR’s practical administration and scoring issues

Another potential obstacle for the use of projectives in the workplace is that of administration and scoring. First, a majority of these measures must be administered individually, which can be expensive and time-consuming. *Reiger* (1949) stated, “Therefore the cost of administering these techniques may make it impractical to use them, except for jobs which involve at least a moderate investment on the part of the employer” (p. 569). However, efforts to create multiple-choice measures (e.g., SORT, CWF) and group administration measures (group Rorschach, SCTs) have shown some promise. We believe more research in HRM is needed to develop projectives that are easily administered in electronic formats (e.g., computer-based testing) or to larger groups of applicants in person.

Additionally, the scoring of projective tests has historically needed the expertise of trained clinical psychologists. However, this is not feasible for all organizations and would likely invite claims of discrimination under the ADA. Thus, we believe that more straightforward scoring schemes should be developed for HR-centric projective tests. Additionally, normative data for these techniques would eliminate any reliance on clinical interpretation. The CS for the Rorschach (*Exner, 1974*; see above) was a good first step; however, there have been criticisms that these norms are based on small samples and “overpathologize” normal individuals (*Lilienfield et al., 2000*). The studies used to develop these norms are unpublished and unavailable to the public. Thus, we call for more comprehensive and transparent normative efforts for projective tests, particularly for objectively scored measures.

8. Suggestions for avoiding philosophical deadlock

By the term, *philosophical deadlock*, we refer to a discipline divided into camps that are locked in a situation wherein the arguments of one camp do not address the concerns of the other. This appears to have occurred repeatedly in the history of projective testing within HR-related areas as well as psychological science in general. Below, we provide several points that we believe can help the field of HRM avoid this deadlock, making reference to some occurrences of these problems found in our literature review.

8.1. Self-report measures are not the "gold standard"

In conducting this review, we noted that a common debate in the field revolves around using relationships between projectives and self-report measures of the same or similar constructs as evidence for the validity and utility of projectives. We feel that this debate is misguided and ignores the basic theory of projection that is used to develop these measures. The TAT and PRB were both refuted based on such evidence. At first glance this may seem a rational position. After all, self-report personality forms are often referred to as “objective” measures, and often possess highly desirable psychometric properties in a statistical sense. However, we argue that the relation of projective and self-report measures is often irrelevant to both the validity and utility of projective measures.

We take a position similar to *Bornstein (2002)*, McClelland et al. (1989), and *Miner (1978)* that this could in fact represent evidence of discriminant validity as long as the projective measure still relates to important outcomes in expected ways. Therefore, we propose that HRM researchers should seek to establish projective measures’ relations with irrefutable outcomes, such as observed (e.g., number of disciplinary actions) or other-inferred behavior (e.g., performance ratings) in high-fidelity or
real settings. We believe relying purely on self-report measures to establish the validity and utility of projectives in some sense misses the point, particularly when there is a motivation to distort responses. This idea is highly related to our next suggestion.

8.2. Measures are not always competing for the construct

As noted above, projectives have been dismissed in the past due to null relations with self-report measures of the same or similar constructs. For example, the TAT was criticized by Murstein and Mathes (1996) for its lack of relation to self-report Big 5 personality traits. However, we agree with other authors (e.g., Bornstein, 2002) who have stated that this reliance on convergent validity based on the traditional multi-trait multi-method approach (i.e., Campbell & Fiske, 1959) is flawed when assessing projective measures. Self-report measures are likely assessing self-attributed traits that are subject to self-presentation motives. Implicit measures aim to assess non-declarative motives that affect behavior directly (McClelland et al., 1989).

Bornstein (2002) refers to a “process dissociation approach” in understanding relations between projective and self-report measures. He argues that these two types of measures should predict convergent behaviors, be moderately correlated, and be affected by different moderating variables. We think it would be wise for HRM researchers to consider this model when developing projective tests. Additionally, the incremental validity of projective measures over other self-report measures is important for future research. To date, there is little research on this topic, with conflicting evidence for each projective test (Lilienfield et al., 2000). Further, HRM researchers could utilize latent variable models to understand whether the two types of measurement occupy unique locations of the construct space. In other words, the two measurement methods do not necessarily need to compete with one another, but may be complimentary, providing equally important information about the construct of interest.

8.3. Focus on whether the construct causes scores

Theoretically rich measurement should be the goal in pursuing endeavors related to projective testing. As noted by Cervone (2010), psychology has gone from measurement theory to psychological theory in reverse, paying more attention to statistical indicators than striving for realist accounts of the causal link between constructs and scores. Statistical techniques can only do so much. Statistical analyses with external variables do help us understand the utility of such measures, but do not truly address the realist idea of validity, whose aim is to show that the attribute of interest is the cause of score variability (Borsboom, 2005). However, we take the position that validity is a precondition of utility, and therefore is a critical consideration. Without truly establishing validity, we cannot realize the full potential of scientific theories for predicting work behaviors and performance.

9. Concluding remarks

In this article, we hope to have provided an in-depth look at the history of projective testing in HR-related areas, and to have evaluated how the debate over projectives has played out in the past. Our aim was not to criticize, but to utilize the past experiences of the fields of psychology and HRM to encourage thoughtful discourse. We believe explorations of projective testing in HRM science are timely, and we are quite excited about this special issue. Being open to new ideas and new methodologies is paramount to the development of the HRM field, and alternatives to self-report measurement tools for explaining and predicting organizational phenomena are greatly needed. We hope the review and ideas presented here have inspired and energized HRM to look to new possibilities with their eyes on the past!

References


