



A framework for developing word fragment completion tasks



Joel Koopman ^{a,*}, Michael Howe ^a, Russell E. Johnson ^a, James A. Tan ^b, Chu-Hsiang Chang ^c

^a Eli Broad College of Business, Michigan State University Department of Management, N475 North Business Complex, East Lansing, MI 48824, United States

^b Department of Management, Herberger Business School, St. Cloud State University, 720 4th Ave. S, Saint Cloud, MN 56301, United States

^c Department of Psychology, Michigan State University, 346 Psychology Building, East Lansing, MI 48824, United States

ARTICLE INFO

Keywords:

Implicit
Indirect measurement
Word fragment completion
Validation

ABSTRACT

To broaden our understanding of the person-based phenomena that impact organizational behavior, researchers are increasingly making use of implicit measures. Explicit measures tap affect, attitudes, and self-concepts that are accessible to introspective awareness, which are sometimes unreliable or distorted by response biases. In contrast, implicit measures assess mental content and processes that operate outside awareness. Unfortunately, the ad hoc manner in which implicit measures are sometimes developed is problematic. As a way of improving research on implicit content and processes, we present a detailed and practical framework for developing one type of implicit measure: word fragment completion tasks. Such tasks have been successfully used to measure a variety of individual difference variables in previous organizational research. Our framework draws on previous research and well understood psychometric principles to describe a process for creating reliable and valid word fragment completion tasks.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The study of attitudes, motivation, traits, and affect is central to research in organizational behavior and human resources management (OBHRM). As early as the 1930s, Allport declared “attitude to be social psychology’s ‘most distinctive and indispensable concept’” (Allport, 1935, as cited in Greenwald & Banaji, 1995, p. 4). Indeed, a variety of work attitudes such as job satisfaction (Judge, Thoresen, Bono, & Patton, 2001) and organizational commitment (Meyer, Stanley, Herscovitch, & Topolnytsky, 2002) have been shown to predict a range of important organizational outcomes (e.g., task performance, contextual performance, and turnover). Similarly, research on motivation enjoys a long and storied research tradition with OBHRM (e.g., Latham & Locke, 2008) and is generally recognized to be critical to both job performance and overall organizational effectiveness (Kanfer, 2009). Personality trait research, though largely seen as of little consequence (at least in HRM research) for a number of years (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007), was wholeheartedly revived owing to research on the applicability of personality to personnel selection (e.g., Barrick & Mount, 1991; Day & Silverman, 1989; Tett, Jackson, & Rothstein, 1991). Finally, affect research has recently gained substantial momentum (e.g., Brief & Weiss, 2002; Fisher & Ashkanasy, 2000) and has been linked with important organizational behaviors both conceptually (e.g., Isen & Baron, 1991; Weiss & Cropanzano, 1996) and empirically (e.g., George, 1991; Kaplan, Bradley, Luchman, & Haynes, 2009; Staw & Barsade, 1993).

Importantly, much of this research within OBHRM has relied primarily on explicit measures, which assess mental content and processes that respondents both consciously have access to and are willing to report (Wittenbrink & Schwarz, 2007). Explicit content and processing is the product of rule-based propositional reasoning in working memory (Smith & DeCoster, 2000; Strack & Deutsch, 2004). There are, however, potential limitations with the use of explicit measures: responses can be unreliable or inaccurate due to faking (e.g., Morgeson et al., 2007), socially desirable self-presentation (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), or self-deceptive behavior (Barrick & Mount, 1996; Gur & Sackeim, 1979). The use of explicit measures may also be limited

* Corresponding author. Tel.: +1 517 884 1674.

E-mail address: koopman@bus.msu.edu (J. Koopman).

because some mental content and processes operate at least partially outside people's awareness and control at implicit levels, yet still exert influence on behavior (Barsade, Ramarajan, & Westen, 2009; Greenwald & Banaji, 1995; Johnson & Steinman, 2009; Johnson & Tan, 2009). Thus, dual processing at explicit and implicit levels may give rise to different sources of affect, attitudes, and self-concepts (Haines & Sumner, 2006). These potential limitations suggest that researchers may be missing something when they rely exclusively on explicit measures (Bing, LeBreton, Davison, Migetz, & James, 2007; Bing, Stewart, Davison, Green, McIntyre & James, 2007; Haines & Sumner, 2006; Johnson & Steinman, 2009). For these reasons, the use of implicit measures is gaining popularity in OBHRM research.

Implicit measures offer a means to assess implicit social cognition, which refers to cognition that is shaped by past experience and affects current processing yet occurs without introspective awareness (Greenwald & Banaji, 1995; Haines & Sumner, 2006). Implicit content and processing involve the accessibility and spreading of activation among concepts in associative memory systems (Smith & DeCoster, 2000; Strack & Deutsch, 2004). Methodologically, technological advancements for assessing implicit social cognition have permitted investigation through the measurement of, for example, response latencies (e.g., Greenwald, McGhee, & Schwartz, 1998) and neural activity (e.g., Ito & Cacioppo, 2007). From a theoretical perspective, attitudes (e.g., Leavitt, Fong, & Greenwald, 2011), motives (e.g., Latham, Stajkovic, & Locke, 2010), traits (e.g., James, 1998), and affect (e.g., Barsade et al., 2009) have been proposed to have an implicit component. By failing to consider the implicit aspects of these phenomena, our understanding is incomplete and their influence on employee behavior is underestimated.

In order to assess implicit social cognition, specific tools must be added to researchers' measurement arsenal. Given the importance of developing and utilizing these new tools, our goal in this paper is to describe a process of developing one type of implicit measure: word fragment completion tasks. These tasks involve presenting participants with a series of word fragments (e.g., "B O _ _ N G") that can be made into one of several words, depending on the letters that are added (e.g., "BORING" or "BOXING"). The words that respondents generate indicate the accessibility or strength of content at implicit levels (e.g., words like "ANXIOUS" and "FEAR" denote negative affectivity). Importantly, this task has recently been successfully used for OBHRM research. For example, Johnson and colleagues have developed word fragment completion measures of trait affectivity (Johnson, Tolentino, Rodopman, & Cho, 2010), self-identity (Johnson & Lord, 2010; Johnson & Saboe, 2011), and approach/avoidance motives (Way & Johnson, 2010). Results from these initial studies indicate that responses on word fragment completion measures correspond to activation at implicit levels, which influences various work outcomes (e.g., task performance and cooperation) both incremental to and in conjunction with processing at explicit levels (Uhlmann et al., 2012).

Though not the only tool available for measuring implicit social cognition (for reviews, see Bing, LeBreton, et al., 2007; Uhlmann et al., 2012), we believe that word fragment completion tasks are an important and underused tool for this type of research. We focus on word fragment completion tasks specifically because they have several advantages compared to other implicit measures. For example, they are easily administered in a paper-and-pencil format unlike other commonly used implicit measures (e.g., Implicit Association Test; Greenwald et al., 1998, although see Lemm, Lane, Sattler, Khan, & Nosek, 2008), and, once developed, they are relatively effortless and inexpensive to administer and score. Word fragment completion tasks can also be adapted to use as part of on-line data collection efforts, such as The Study Response Project (<http://www.studyresponse.net>) and Mechanical Turk (Buhrmester, Kwang, & Gosling, 2011). Another advantage is that word fragment completion tasks are quite flexible—they can be used to measure the accessibility of implicit content or, when paired with priming techniques, used to measure implicit associations (Uhlmann et al., 2012). For example, Anderson and colleagues (Anderson, Carnagey, & Eubanks, 2003; Anderson et al., 2004) found that participants exposed to violent media lyrics were more likely to generate aggressive words. In addition, priming effects on word fragment completion tasks have been used to study attitudes, prejudice, and stereotyping (e.g., Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Gilbert & Hixon, 1991; Sinclair & Kunda, 1999). Thus, although word fragments have been used to assess personal characteristics for OBHRM research, when combined with priming, it is possible to utilize such tasks to assess associative knowledge like attitudes and stereotypes. A final advantage is their novelty—participants often comment how these 'word games' help break up the monotony of responding to survey item after survey item. If so, then this enjoyment may enhance respondent engagement and attention when completing studies, thereby enhancing the quality of their data.

1.1. Implicit versus explicit measures

Before we provide recommendations for how to construct and validate word fragment completion tasks, we first discuss our view of implicit versus explicit measures. Some scholars have suggested that implicit and explicit measures offer alternative ways of measuring the same underlying construct (e.g., Vargas, Sekaquaptewa, & von Hippel, 2007), albeit they may reflect uncorrelated components of the same construct (e.g., Bing, LeBreton, et al., 2007) or the same construct at different processing stages (Fazio & Olson, 2003). Alternatively, other scholars propose that implicit and explicit scores reflect distinct yet related constructs (e.g., McClelland, Koestner, & Weinberger, 1989; Nosek & Smyth, 2007). Although we make no attempt to resolve this debate in the current paper, dual processing theories indicate that there will often be some overlap among implicit and explicit scores due to spillover between rule-based processing at explicit levels and associative processing at implicit levels (Smith & DeCoster, 2000; Strack & Deutsch, 2004). Consistent with this idea, small to moderate positive correlations have been observed among explicit and implicit scores (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, 2005; Nosek & Smyth, 2007). Thus, a few of our recommendations assume that there exists some overlap among explicit and implicit scores, although we realize this assumption will not be true in every case. Ultimately, the nature of relations between implicit and explicit scores is a

question that requires both theory and empirical research to answer, and researchers should clearly state their assumptions regarding this relation.

Fig. 1 represents our view of how explicit and implicit measures each capture unique and shared variance in underlying constructs. Both types of measures assess the same phenomenon (albeit at different levels of information processing), and do so with some amount of both contamination (regions A, F, and D) and deficiency (i.e., region G), consistent with Cook and Campbell's (1979) notion of mono-method bias. In addition to explaining unique variance in the focal construct (i.e., regions B and C for implicit and explicit measures, respectively), these measures also exhibit some overlap. Some of the overlap may reflect a portion of the focal construct that is captured by both explicit and implicit measures (i.e., region E), while others may reflect a shared contamination between the two types of measures (i.e., region F). Take, for example, the assessment of negative affectivity using explicit and implicit measures (Johnson et al., 2010). Scores on a word fragment completion task in part reflect the accessibility of negative affectivity at implicit levels (region B), and irrelevant variance due to verbal ability (region A). Scores on an explicit measure capture both propositional reasoning at explicit levels (e.g., "I felt fatigued and was impatient with my spouse this morning, therefore I must be in a negative mood;" region C) as well as error variance due to response biases (region D). As mentioned above, we also expect some overlap between explicit and implicit scores (regions E and F).

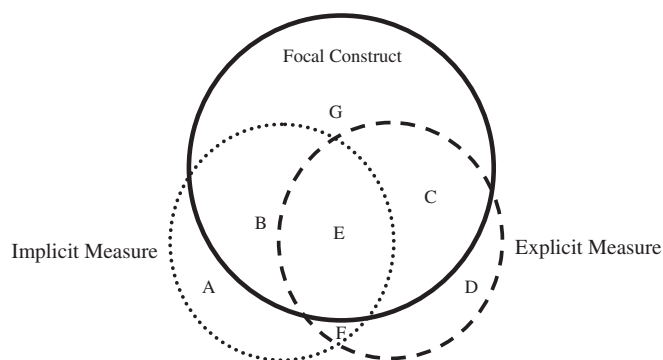
Of course, the proportions of the various regions in Fig. 1 vary depending on the nature of the focal construct, the specific measures used, and characteristics of respondents (e.g., Nosek, 2005). Although meta-analytic evidence supports that there is at least some overlap between explicit and implicit measures (e.g., Hofmann et al., 2005), it is possible that they tap unique components of the focal construct (i.e., nonexistent E and F regions; Bing, LeBreton, et al., 2007; Bing, Stewart, et al., 2007). In the case where the distinctiveness of explicit and implicit measures is theoretical (Bing, LeBreton, et al., 2007) and cannot be attributed to their poor psychometric properties (Lilienfeld, Wood, & Garb, 2000), we believe that it is beneficial to utilize both types of measures, as they provide complementary yet non-redundant information. With these assumptions in mind, in the next section we present guidelines and recommendations for constructing and validating word fragment completion tasks.

2. Framework for constructing word fragment completion tasks

Although word fragment completion tasks have been used in psychological research for nearly 30 years (e.g., Tulving, Schacter, & Stark, 1982), appropriate methods for their construction, validation, and administration have not been specified. We believe that this is an oversight, as theoretical progress in a field depends strongly upon the methods and measures used to test theories (Edwards, 2008), and poorly developed measures impair our ability to understand organizationally relevant phenomena (Hinkin, 1995). Thus, the aim of this paper is to provide a framework to guide researchers when developing and administering word fragment completion tasks. These suggestions and recommendations are applicable regardless of whether or not word fragment items are used in conjunction with priming techniques. Outlined in Fig. 2 are six steps for developing and administering word fragment completion tasks, which are adapted from existing practices for explicit measures (e.g., Hinkin, 1998).

2.1. Item creation

Our recommendations begin in much the same place as most OBHRM research—with theory. The decision to use an implicit measure is a theoretical argument by researchers that the focal construct is not adequately captured by explicit measures



Note: Region A: Unique implicit measure contamination; Region B: Unique implicit measure construct validity; Region C: Unique explicit measure construct validity; Region D: Unique explicit measure contamination; Region E: Implicit and explicit measure common construct validity; Region F: Implicit and explicit measure common contamination. Regions C+G: Implicit measure deficiency; Regions B+G: Explicit measure deficiency.

Fig. 1. Conceptual model relating explicit and implicit measures.

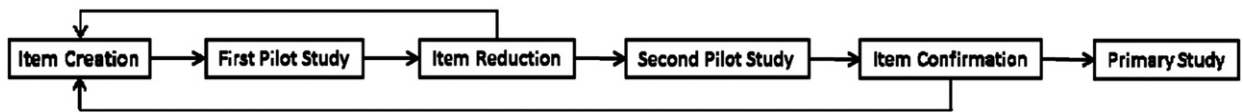


Fig. 2. Steps for developing a word fragment completion task.

(whether for faking, self-presentation, or other reasons) or that in some aspect the construct operates outside awareness and control. For example, fundamental psychological needs, motives, values, and traits are believed to operate at implicit levels (Strack & Deutsch, 2004), thus the use of implicit measures are appropriate when examining basic phenomena like approach and avoidance motivations (Elliot, 2006) and positive and negative affectivity (Johnson et al., 2010). In the interest of parsimony, there is little reason to utilize implicit measures if explicit measures are sufficient for addressing research questions. If, however, there is a theory supporting the need for assessment at implicit levels, then researchers should proceed by developing an appropriate measure.

Once the decision to employ a word fragment completion task has been made, the next step involves generating a list of words pertaining to the focal variable. One method is to use words from validated scales. For example, emotion words (e.g., joy, afraid) taken from explicit measures of trait affectivity (Watson, Clark, & Tellegen's, 1988, Positive and Negative Affectivity Schedule) served as some of the target words on the implicit measure used by Johnson et al. (2010). Another approach is to recruit people from the to-be-sampled population to generate words that relate to the focal construct. For example, Steele and Aronson (1995) created an implicit measure of stereotype activation by using words generated by undergraduate students that reflected stereotypic beliefs about African Americans. An advantage of this approach is that the target words are more common and meaningful to respondents when the samples used for item creation and test administration are drawn from the same population. A third approach is to have subject matter experts, including the researchers, generate target words. For example, Johnson and colleagues (Johnson & Lord, 2010; Johnson & Saboe, 2011) created their own sets of independent (e.g., "me" and "self") and interdependent (e.g., "we" and "team") words to use as targets on their word fragment completion measure of self-identity.

In order to adequately sample the content domain of the construct, a large number of potential words should be generated. Many of the original words are likely to be lost in the item reduction phase, so generating a large initial list ensures that a sufficient number of words are retained for the final measure. In the case of explicit measures, it is expected that 50% of the initial items will be lost through the scale development process (Hinkin, 1998). This number is likely higher for word fragment items because of word frequency issues. In our experience, several word fragments are subsequently dropped or changed due to a lack of variance in responses. Note, however, that there is flexibility in creating word fragments because different configurations are possible for the same target word. For example, the fragments "S _ _" and "_ A D" are two possibilities for the target word "SAD." Thus, an efficient approach is to create multiple fragments for the same word and retain the one that produces a sufficient degree of variance in responses.

To determine the number of initial items to create, a target for the number of desired items must first be established. The number of items has varied across studies. For example, Johnson et al. (2010) used 20 items to assess positive and negative affectivity and Johnson and Lord (2010) used 25 items to assess independent and interdependent self-identity levels. Alternatively, to assess stereotype activation, Son Hing, Li, and Zanna (2002) used 5 items and Sinclair and Kunda (1999) used 13. Our general guidance for scale length follows that for explicit measures. Namely, enough items must be retained to adequately measure the construct without unnecessarily burdening respondents (Schmitt & Stults, 1986). Additionally, given that coding responses on word fragment completion measures require additional time and multiple coders, it is important to avoid unnecessarily long scales that burden researchers as well (Carmines & Zeller, 1979). Although it is acceptable to create explicit measures with 4–6 items (Hinkin, 1998), this number is too few for word fragment completion tasks because respondents may not generate a word for every item and not all words will correspond to the focal construct. For example, Johnson and Saboe (2011) found that on average participants generated a word for 75% of the items and 38% of responses corresponded to target words. In order to ensure there is sufficient opportunity to assess the focal construct, we recommend using 12–15 word fragments per construct. Doing so means that most participants would generate at least 10 words, which provides a sufficient sample of responses to reliably evaluate their standing on the focal construct.

After the list of potential target words is generated, the content validity of the items should be assessed. The technique of substantive validity analysis (Anderson & Gerbing, 1991) is appropriate for this purpose. For this technique, a small sample provides ratings of whether each word fits with the definition of the focal construct, allowing for the creation of two validity indices. The first index, termed the *proportion of substantive agreement*, represents the proportion of raters assigning each word to its intended category. We recommend retaining words that are correctly assigned by at least 80% of raters. The second index, termed the *substantive-validity coefficient*, is calculated by taking the difference between the number of raters who assigned a word to its intended category and the number of raters who assigned the word to the next most popular category. This difference score is then divided by the total number of raters. Importantly, a statistical significance test is available for the substantive-validity coefficient, with the null hypothesis being that the calculated proportion is not different from .50 (the expected value if the word is equally represented in both categories). Rejecting the null hypothesis, combined with a sufficiently high proportion of substantive agreement, is taken as evidence for the content validity of the words.

After the list of validated target words is finalized, the next step is to create word fragments by removing one or more letters from each word. Deciding how many letters to remove, and which letters to remove, is important because it impacts respondents'

degrees of freedom for generating different words. The goal is to ensure sufficient variance in responses, thus avoiding the creation of items that produce too high of a response rate for both target and non-target words as well as items that require substantial cognitive deliberation to provide a response. The expectation is that participants with high levels on the focal construct are more likely to generate words that reflect that construct when presented with a word fragment cue. It is important to create fragments such that it is relatively easy for respondents to generate a word. If a response is not immediate but rather the result of a deliberative trial-and-error approach or memory search, then responses are more indicative of participants' vocabulary rather than the accessibility of the target content at implicit levels.

The difficulty of generating a word is dependent, in part, on the location of the missing letters. Typically, it is easier to generate words (e.g., "GROUP") when letters are removed from the end of the word (e.g., "G R _ _") versus the beginning (e.g., "_ _ _ U P") or middle (e.g., "G _ _ _ P"). Thus, the location of blanks can be strategically placed to increase the difficulty of items with too many target responses or increase the ease of items with too few target responses. Furthermore, the more blanks that are used, the more degrees of freedom respondents have to create words. It is important to have variance in the responses, however too many blanks may lead a respondent to engage in effortful cognitive processing. In general, fewer blanks are better (e.g., "TEAM" may be assessed with "T E A _" or "T E _ _", however "T _ _ _" may result in low frequency of the target word). As mentioned earlier, an efficient approach is to create two or more fragments for each word, and then retain whichever item has the optimal degrees of difficulty and variance in responses based on pilot data. In order to help identify possible words that can be created from various word fragments with one or more missing letters, online crossword solution tools can be utilized (e.g., <http://www.crosswordsolver.org>). Tools such as these may detect additional target words that were not thought of when items were first created.

Researchers may also consult dictionaries of word frequencies to ensure that target and non-target words have comparable frequencies. Sources such as the English Lexicon Project (Balota et al., 2007), the *Frequency Dictionary of Contemporary American English* (Davies & Gardner, 2010), and the University of South Florida Free Association, Rhyme, and Word Fragment Norms (Nelson, McEvoy, & Schreiber, 2004) indicate how common words are in their everyday use. If there are large discrepancies in frequencies for an item (i.e., target words have particularly low frequencies and/or non-target words have especially high frequencies), then it is a poor item as there will be a little variance in responses. Each target word should have comparable frequencies with other possible (neutral) words to maximize the variance in responses.

2.2. First pilot study

Once a sufficient number of potential word fragment items have been generated, the next step is to pilot test them. Pilot testing is essential because it provides an opportunity to evaluate and refine the measure as well as uncover and fix potential issues before the expense and effort of a full scale administration is undertaken (Nunnally & Bernstein, 1994). In actuality, we recommend the use of two pilot tests: one for initial item analyses and a second for assessing construct validity. In this section we discuss issues concerning the administration and scoring of word fragment completion tasks. The following section ('Item reduction') describes analyses to be conducted on the data from the first pilot study.

Because word fragment completion tasks are intended to assess content outside respondents' awareness and control, it is necessary to minimize prolonged deliberation by respondents when completing the items (Vargas et al., 2007). Thus, instructions should clearly indicate that responses must be given quickly, that the respondents should neither deliberate over an item nor change their response once they have provided it, and that skipping items when a word does not immediately come to mind is acceptable. Specific time limits can also be used to ensure that respondents complete the task quickly. When possible, it is good practice to proctor the administration of paper-and-pencil versions of word fragment completion tasks to ensure that respondents follow instructions. When it is not possible to proctor the survey, another option is to include one or several items at the end of the instrument that asks respondents questions such as "how long did you deliberate over the items" or "did you go back and change items after you had answered." Such items can be used to identify and possibly disqualify individuals who did not appropriately follow instructions. An advantage of computerized administration of word fragment completion tasks is that a fixed time limit (e.g., 5 seconds) to complete each word fragment can be set up. Once the time expires, the next word fragment is presented. Alternatively, the amount of time respondents take to complete each item can be recorded and used for data screening purposes (e.g., responses that exceed a maximum time limit are discarded). Precautions such as these are helpful for ensuring that responses reflect the most accessible content at implicit levels.

A second issue to consider is the possibility that participants' responses on earlier word fragments may prime their responses on subsequent word fragments, which is also a concern for explicit measures (Feldman & Lynch, 1988). For example, providing the response "HAPPY" to the fragment "_ _ _ P Y" may in fact prime respondents to generate the word "JOY" when completing the subsequent fragment "J _ _". Such self-generated validity effects occur independent of people's actual standing on the construct of interest. One way to minimize this possibility is to include filler items among the focal word fragments. After the initial pool of items is reduced following the first pilot study, adding a handful of filler items should not make the length of the word fragment completion task prohibitive.

Another issue pertains to the similarity of the conditions for pilot testing and actual testing. Ideally, the pilot test administration and sample should match the planned method and sample for the main study (Nunnally & Bernstein, 1994). For example, if the primary study involves undergraduates in a laboratory setting with an experimenter present, then the pilot study should be conducted as such. Alternatively, if the survey will be administered online, then the pilot study should also follow this format. The

equivalence of word fragment completion tasks administered via paper-and-pencil or via computer has yet to be assessed, and until it is, the format of the task should be held constant across the pilot and primary studies when possible.

Researchers should also consider the use of different experimental conditions during pilot testing, depending on the nature of the construct. Many attitudes and individual differences can be manipulated (e.g., satisfaction and intrinsic motivation), and experimental manipulation is a particularly useful tool for assessing the validity of implicit measures. For example, when developing their word fragment measure of self-identity levels, [Johnson and Saboe \(2011\)](#) pilot tested their measure by using established vignettes that primed either an independent or interdependent identity. Following the manipulation, the word fragments were administered to the participants. By examining participants' responses, the authors were able to identify what items best reflected implicit identity activation and best discriminated between the experimental conditions. Including a control (non-primed) condition is also useful because it provides insight into the natural frequency and type of words that participants generate. Given that much OBHRM research is non-experimental, normative data is informative for interpreting survey results. Note also that state (versus trait) phenomena are more responsive to experimental manipulations and thus, other validation procedures besides priming techniques may be needed when developing word fragment items that are intended to tap stable dispositions.

A final issue to tackle during pilot testing involves scoring the word fragment completion task. Scoring is a two-step process: responses must first be coded, and then an overall score must be derived. At least two subject matter experts, preferably ones not directly associated with the study, should be used for coding. If necessary, coders should receive training on the focal construct. The job of coders is to independently rate whether each word generated by participants reflects the focal construct or a neutral concept (there will also be missing cases when items are left blank). Although coders will have a list of a priori target words from the item creation stage, participants may generate additional target words for an item. For example, to measure positive affect researchers might use the word fragment "S T R _ _" for the target word "STRONG." However, if a respondent were to give "STRAIN," this could be interpreted as an indicator of negative affect. Thus, coding should not be limited merely to a priori target words, at least during the initial phases of scale development. Given that the initial list of target words will evolve over time, it is important to assess initial inter-rater agreement (e.g., Cohen's kappa [[Cohen, 1960](#)] when there are two reviewers). Any discrepancies in coding can then be resolved through discussion with the principal investigator.

Two other coding issues warrant mention. First, a decision must be made regarding how to handle words that reflect the focal construct yet are either misspelled (e.g., "ENVYOUS" instead of "ENVIUOUS") or whose letters exceed the number of blank spaces (e.g., "ANGER" as a response to "_ _ E R"). Options include (a) being open to considering the response as a target word if appropriate, or else coding the response (b) as a neutral word or (c) as missing. Unfortunately, there is no research on the ramifications of spelling mistakes or length violations on the validity of word fragment items. Until there is, our recommendation is to select one of the above options and then consistently apply it. The second issue is a question of missing data—namely, how many skipped items are too many to justify retaining a participant's data? Recall that participants are instructed to skip items when a word does not immediately come to mind in order to ensure that their responses reflect the most accessible content at implicit levels. While it is certainly possible that a participant may skip the majority of items, this runs counter to the available empirical information. [Johnson and Saboe \(2011\)](#) observed that, on average, participants completed 75% of the word fragments and no participants completed fewer than 60% of the items. As a practical recommendation, participants ought to have completed a sufficient number of items—at least 50 or 60% depending on scale length—to provide a reliable estimate of the person's standing on the focal construct.

The second step of scoring involves deriving an overall score after all responses have been coded. One of two approaches are typically used: a total score based on the sum of the target words (e.g., [Sinclair & Kunda, 1999](#)) or proportion score where the sum of the target words are divided by the total number of words generated (e.g., [Johnson & Lord, 2010](#)). Our recommendation is to use the latter approach because the variance in the number of items completed across participants can be quite high (recall that participants are instructed to skip items when a word does not immediately come to mind in order to ensure that responses reflect the most accessible content at implicit levels). Suppose, for example, that two participants each generated four target words on a ten-item word fragment completion measure. If one participant completed a total of five items (i.e., 80% hit rate) while the other completed all ten items (i.e., a 40% hit rate), the focal construct is clearly more accessible for the first participant. Although there are some drawbacks with using proportion scores ([Cohen, Cohen, West, & Aiken, 2003](#)), and researchers should be aware of the assumptions and limitations of their use, such scores are more appropriate than raw sums when scoring word fragment completion measures.

2.3. Item reduction

Once the measure has been piloted, the next step is to identify poor-quality items, such as those that fail to exhibit sufficient variance or that do not relate to manipulations or other measures of the focal construct. In a neutral or control condition, each item should result in a target word between 25 and 75% of responses as compared to non-target words. Ideally, each item would result in a target word 50% of the time because, as the proportion of target to non-target words for each item becomes more extreme, the maximum attainable correlation between that item and other variables becomes progressively smaller. Our recommended 25-to-75% range is based on existing guidelines for subgroup proportions ([Cohen et al., 2003](#)). Unfortunately, most studies do not report variance ranges (one exception is [Johnson & Saboe, 2011](#), who used a range of 15–85%), making it difficult to judge what the common practice is. In a manipulated condition, the percentage of target words should be more extreme and

significantly different from a control or alternative condition. Items that fail to perform as expected should be discarded or revised (and then re-evaluated).

The quality of items can also be assessed via a series of inter-item analyses. One option is to calculate item inter-correlations, and then eliminate items that correlate weakly (e.g., $r_s < .40$) with the other items (Kim & Mueller, 1978). Although the set of items might also be evaluated by calculating item-to-total correlations, in the following section we argue that test–retest reliability is more appropriate than internal consistency when assessing the reliability of word fragment completion tasks. Items with low inter-item correlations add error into the measure and may not be adequately tapping the focal construct (Churchill, 1979). In the case of the word fragment items, because items are dichotomous (e.g., responses are coded as a “1” if it corresponds to the target construct or “0” if it is a neutral word), phi coefficients should be calculated (Anastasi & Urbina, 1997); these coefficients are interpreted similarly to a Pearson zero-order correlation.

After word fragments with weak inter-item correlations are removed, exploratory factor analysis (EFA) can be used to provide additional evidence of the scale's construct validity (Guadagnoli & Velicer, 1988). For example, Johnson and Saboe (2011) utilized EFA to examine the extent to which items that assessed independent and interdependent self-identities loaded on their appropriate factor. The authors discarded items with weak loadings on the primary factor and those with high cross-loadings on the other factor. Johnson and Saboe (2011) used a cutoff value of .25 for the loading; generally .40 is the rule of thumb for EFA factor loadings (Stevens, 2002). Given the exploratory nature of this analysis, as well as the open-ended nature of word fragment items, a factor loading cutoff of .30 is acceptable. Alternatively, Hinkin (1998) suggests that an item may be kept if its loading on the appropriate factor is twice as large as its loading on any other factor. Regardless, we encourage researchers to be consistent in their application of criteria and to report these values in their manuscript.

2.4. Second pilot study

Upon conclusion of the item reduction phase, a potentially viable word fragment completion task exists. We emphasize “potentially” as the measure has yet to be subjected to confirmatory analyses to demonstrate that it indeed taps a unique portion of the underlying construct as it is designed to do. As with the first pilot test, this test should be administered in a similar setting as the planned primary study. The instructions should be similar (e.g., skipping items is acceptable, complete the items as quickly as possible, and so forth).

Whereas the purpose of the first pilot study was to develop items, the purpose of this pilot study is to provide evidence of its reliability and validity. Below we suggest that test–retest reliability is the most appropriate index of reliability for word fragment completion tasks, which requires that participants for this pilot sample should be recruited in such a way that they may be available again for a second administration. The validity of word fragment completion tasks may be assessed by analyzing their convergence with other previously validated measures of the focal construct as well as their relationships with measures of other constructs known to be in the nomological net of the focal construct. For example, an implicit measure of extraversion should correlate with scores on explicit measures of the same construct¹ as well as scores on implicit measures of related constructs, such as positive affectivity (both extraversion and positive affectivity are markers of approach motivation; Elliot, 2006). As such, multiple measures should be administered as part of the second pilot test. It is good practice to counterbalance the ordering of the implicit measure and other validated measures of the focal construct and then test to see whether responses on the implicit measure are altered as a result. It is possible that order effects exist for implicit measures because they are susceptible to priming (Ferguson & Bargh, 2007; Smith & Conrey, 2007; Wittenbrink & Schwarz, 2007).

2.5. Item confirmation

2.5.1. Reliability

After the word fragment completion task has been administered twice to the pilot sample, the test–retest reliability of the measure can be assessed. This is an important and often overlooked aspect of developing implicit measures. Indeed, many researchers do not discuss reliability at all. We advocate the use of test–retest reliability for word fragment completion tasks. We recommend this method as opposed to the internal consistency, parallel forms, and split-half approaches for several reasons. First, the test–retest method provides a middle-of-the-road estimate of reliability as opposed to an upper or lower bound estimate. The split-half approach tends to overestimate reliability whereas parallel forms tend to underestimate it (Gatewood & Feild, 2001). The split-half approach is furthermore problematic due to the amount of skipped items in the measures as well as the generation of non-target words. Skipped items on word fragment measures are high because of instructions to skip items for which a response does not immediately spring to mind, and even when items are completed, a large percentage will be non-target words. Finally, given the nature of the task (e.g., come up with words as fast as possible), word fragment completion tasks are arguably “speed tests” in nature. As such, the use of single-trial reliability coefficients (e.g., split-half) is inappropriate (Anastasi & Urbina, 1997; Nunnally & Bernstein, 1994). Lastly, internal consistency is not viable due to the open-ended nature of the items. On an explicit measure, the semantic meaning of the items is derived from the exact wording of the items—that is, participant responses are the result of item meaning. Thus, items are interpreted consistently across people. Word fragment completion tasks, however, are akin to projective measures in that item meaning derives from participant responses—that is, participant responses give

¹ As mentioned earlier, we assume that there is at least some degree of overlap between the implicit and explicit measures. In the case where the two types of measures are unrelated to each other yet still reflect the underlying construct (e.g., Bing, LeBreton et al., 2007), this guideline is not applicable.

meaning to the item. Therefore, it is difficult to identify a set of responses that are internally consistent because item meaning can vary across respondents.

For the reasons above, test–retest is the preferred method to evaluate the reliability of word fragment completion tasks. For relatively stable constructs (e.g., trait affectivity), a 5–6 week window is appropriate for administration of the second test. However, this window should be shortened (e.g., to 1 or 2 weeks) for less stable constructs (e.g., work engagement; [Christian, Garza, & Slaughter, 2011](#)) given that rank-order changes are expected for such constructs.

2.5.2. Validity

Once the reliability of the measure has been assessed, a series of confirmatory factor analyses (CFA) should be conducted to verify the factor structure and significance of individual item loadings. We recommend CFA as opposed to EFA here for three reasons. First, CFA provides quantifiable fit statistics regarding the hypothesized factor structure ([Long, 1983](#)). Second, CFA provides a stronger test of the a priori hypothesis that an underlying construct is the cause of the observed responses on the word fragment items. Finally, CFA provides the opportunity to assess many of the validity considerations that are described below (e.g., testing a measurement model that includes multiple measures of the focal construct and other constructs in the nomological net).

The measurement model estimates relations of the implicit measure with previously validated measures of the same construct as well as measures of related constructs known to be in the nomological net of the focal construct ([Cronbach & Meehl, 1955](#)). As mentioned earlier, it is expected that small to moderate correlations will be observed between implicit and explicit scores (see [Fig. 1](#)). Note, however, that the degree of convergence between implicit and explicit scores is moderated by the focal construct. Convergence will be lower for constructs that involve self-evaluations (e.g., self-esteem) and socially sensitive information (racial or gender stereotypes) due to socially desirable responding and impression management biases that alter scores on explicit measures ([Fazio & Olson, 2003](#)). For less threatening and socially sensitive constructs, such as trait affectivity and self-identity, moderate correlations have been observed between implicit and explicit measures (e.g., [Johnson & Saboe, 2011](#); [Johnson et al., 2010](#)).

At this stage of measure development, discriminant validity can also be assessed by including measures that have no theoretical connection with the focal measure. For example, a word fragment completion measure that assesses affectivity should not be related to constructs such as conscientiousness (e.g., [Johnson et al., 2010](#)) or cognitive ability (e.g., [Stokes & Levin, 1990](#)). Though it is possible that implicit and explicit measures are uncorrelated due to the nature of the construct (e.g., dual process models of behavior; [McClelland et al., 1989](#)) or how the measures are constructed (e.g., [Bing, LeBreton et al., 2007](#)), the lack of any association is also likely the result of measurement error or a lack of conceptual correspondence between the explicit and implicit variables ([Gawronski, 2009](#)). Thus, we caution researchers that the lack of a relationship between implicit and explicit measures should not be taken automatically as evidence for discriminant validity. Rather, theoretical rationale for the expected relationships or distinctiveness between the two measures is necessary before such evidence can be properly evaluated. We would also caution against adding too many measures that may not be conceptually related to the focal variable as this may lead to questionnaire fatigue which may also introduce the construct-irrelevant variance.

2.6. Primary study

After verifying the reliability and validity of the word fragment completion measure, it is now ready to use it for hypothesis testing. When conducting the primary study, as in the second pilot study, we encourage researchers to include an explicit measure of the focal construct in addition to the implicit measure. We encourage doing so because it enables researchers to assess the incremental validity of the word fragment completion task relative to an explicit measure. Incremental validity can be assessed in a number of ways ([LeBreton, Hargis, Griepentrog, Oswald, & Ployhart, 2007](#)). For example, by using the hierarchical regression, the implicit score can be added in a subsequent step following the explicit score. The addition of the implicit measure should result in a significant ΔR^2 . Analogous to this approach is specifying a structural equations model that includes both measures in the model; a significant path from the implicit score to the criterion indicates that it predicts incremental to the explicit score. Perhaps more appropriately, researchers should assess the relative importance of the implicit measure in predicting the outcome of interest by using dominance or relative weights analysis ([LeBreton et al., 2007](#); see also [Johnson & Saboe, 2011](#); [Johnson et al., 2010](#)). These analyses indicate how much observed variance in the criterion is attributable to the implicit and explicit scores. Importantly, these are only suggestions as to how predictive validity may be demonstrated; researchers may choose alternative methods as well. Beyond demonstrating predictive validity, another advantage of including both implicit and explicit measures is that integrative models (e.g., mediated and moderated effects) can also be tested ([Bing, LeBreton, et al., 2007](#); [Bing, Stewart, et al., 2007](#); [Winter, John, Stewart, Klohnen, & Duncan, 1998](#)).

3. Discussion

We agree with [Hinkin \(1998, p.104\)](#) that “the adequate measurement of abstract constructs is perhaps the greatest challenge to understanding the behavior of people in organizations.” As such, we have laid out a methodology for developing effective word fragment completion tasks to help in this endeavor when researchers are interested in examining implicit content and processes. Beyond simply aiding in developing effective word fragment completion measures, we also hope that our recommendations will embolden researchers to consider the role played by implicit content and processes in their research. Although implicit phenomena are evident in the roots of psychology (e.g., Freud’s treatment of the unconscious), subsequent research on organizational psychology has primarily emphasized attitude, motivation, traits and affect at explicit levels. We believe that there

is more depth and complexity to the human experience than can be captured solely by explicit measures, and important contributions can be made by considering phenomena that occur outside people's awareness and control.

It is important, however, to bear in mind that OBHRM is a field that has long been dominated by investigations of what we can see, or at least what we can describe. Though we have progressed beyond the positivist paradigm that reified observable behavior at the expense of research on traits and attitudes (e.g., Schmidt, Hunter, & Pearlman, 1981), we still tend to rely on individuals' conscious appraisal and reporting of their own attitudes, motivations, traits and affect in our research. An analysis of implicit processes falls outside of this self-report paradigm. For this reason, we suggest that researchers act in the Kuhnian tradition (Kuhn, 1963) of slowly pushing the bounds of our theories and paradigms until they can be pushed no more. Until such a time, the study of the implicit, especially through the means currently available, may meet with hesitance from reviewers, editors, and the scientific community at large. This hesitance, however, is likely to be diminished to the extent that exigent theory indicates the importance of implicit processes in the phenomenon being investigated and further, that theoretically motivated and methodologically rigorous measures lend credence to the proposed relationships.

3.1. Limitations and outstanding issues

In the remainder of this section we address a few outstanding issues regarding the development and use of word fragment completion tasks. The first issue, which applies to all instruments, is that there is no such thing as a "process pure" measure. That is, scores on explicit and implicit measures are both influenced by content and processes at explicit and implicit levels (Gawronski, Lebel, & Peters, 2007). However, implicit measures like word fragment completion tasks capture a larger proportion of variance in content and processes that occur outside awareness and control compared to explicit measures. Developing word fragment completion measures based on the steps listed in Table 1 will produce valid tools for assessing focal constructs at implicit levels.

The second issue concerns the practicality of developing word fragment completion measures. Although our goal was to outline recommended steps in order to maximize the validity of such measures, the numerous steps and the effort required to accomplish each one may appear daunting to many researchers and practitioners. Admittedly, our recommendations represent ideal circumstances when sufficient time and resources are available. In situations where time and resources are insufficient, it may be possible to circumvent some steps yet still develop measures that have acceptable validity. For example, assessing the content validity of target words can be bypassed if the words are pulled directly from existing measures. Likewise, an acceptable degree of validity might be maintained by gathering evidence via one or two methods (e.g., experimental or known group designs) while excluding other methods (e.g., testing the nomological network). Another possibility for streamlining the development and validation process is to combine the two pilot studies into one and run item reduction and item confirmation analyses simultaneously. Where possible, though, it is recommended that most or all of the steps in Table 1 are followed.

The third and final issue pertains to the flexibility of word fragment completion tasks. While many of the examples discussed in this paper have involved personality and self-concept constructs (e.g., trait affectivity and self-identity), these served as examples and by no means are we implying that word fragment completion tasks are limited to measuring such phenomena. Basic needs and motives (e.g., achievement, power, and belonging) are possible candidates, as are attitudes and stereotypes when word fragment completion tasks are paired with priming methods. For example, supervisor satisfaction might be measured by

Table 1
Summary of scale development steps and recommendations.

		Development stage				
		Item creation	First pilot test	Item reduction	Second pilot test and item confirmation	Primary study
Deliverables	An expansive list of potential word fragments	Data related to response frequencies and word scoring guide	Potential measure (12–15 items)	Reliable and valid measure(12–15 items)	Hypothesis testing	
Guidelines	Theoretical justification for examining implicit content	Participants and test conditions similar to those of primary study	Drop items with target responses that exceed 75% or fall below 25%	Participants and test conditions similar to those of primary study	Include direct and indirect measures of focal construct	
	Generate words using subject matter experts and existing direct measures	Use of experimental manipulations for validation purposes	Run EFA and drop items with low primary factor loadings or high cross-loadings	Run CFA to verify factor structure	Counter-balance direct and indirect measures	
	Generate at least twice as many words as needed	Participants instructed to complete items quickly	Drop items with low inter-item correlations	Administer test multiple times to assess test–retest reliability	Use multiple raters who are blind to hypotheses	
	Create multiple fragments for each word	Use multiple raters to code responses and calculate inter-rater agreement		When possible, assess convergent validity by including multiple measures of the focal construct	Assess incremental prediction of word fragment measure	
	Assess content validity of words	Create total score after controlling for number of words generated		Assess criterion-related validity by including measures of related constructs in nomological net	Assess relative importance of word fragment measure	

priming employees to think about their work supervisor and then assessing the number of positive or negative emotion words that they generate. While it is beyond the scope of this paper to identify what constructs can and cannot be assessed (this question is, ultimately, an empirical one), our key advice is to provide reasonable support for the idea that the focal construct operates at implicit levels.

On a related note, word fragment completion measures are also flexible in that they have been used to measure both states and traits. For example, Johnson and colleagues have used their word fragment completion measure of self-identity to measure chronic levels of identity (Johnson & Saboe, 2011) as well as the state-based identity activation in response to unfairness (Johnson & Lord, 2010). Although implicit measures were originally believed to capture more stable mental representations and associations, they are subject to context effects the same as explicit measures (Ferguson & Bargh, 2007; Smith & Conrey, 2007). It is necessary, then, to specify whether the goal is to assess a state or trait and then use appropriate procedures when developing and administering word fragment completion measures. This is especially true when assessing trait or chronic levels of a construct because any affect, motives, needs, etc., that are currently accessible when completing word fragments may bias responses. Thus, steps ought to be taken to reduce any context-based priming effects that respondents may have been exposed to prior to being administered the word fragment completion task. One way might be for respondents to complete a short relaxation exercise or a task that is unrelated to the focal construct (e.g., solving logic or math problems). If the goal is to assess trait levels, it is also important to verify that the word fragment completion measure has sufficient test–retest reliability and shows convergent validity with other measures that tap chronic levels of the same construct. In the case of state levels, it must be shown that priming the focal construct affects the types of words that respondents generate. The take-away point is that different methods for developing and administering implicit measures are needed depending on whether the goal is to assess trait or state levels of the focal construct.

Nevertheless, whether the goal is to measure trait or state levels of phenomena believed to operate at implicit levels, we believe that word fragment completion tasks are a useful tool for capturing such phenomena. The aim of this paper was to specify steps for developing and validating word fragment completion tasks that can be used in laboratory and field settings. Our hope is that making such tasks more accessible will increase the attention that researchers and practitioners pay to implicit content and processes, ultimately enhancing our understanding of affect and cognition in OBHRM.

References

- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Anderson, C. A., Carnagey, N. L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology, 84*, 960–971.
- Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin, A. J., Eubanks, J., & Valentine, J. C. (2004). Violent video games: Specific effects of violent content on aggressive thoughts and behavior. *Advances in Experimental Social Psychology, 36*, 199–249.
- Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology, 76*, 732–740.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.
- Barrick, M. R., & Mount, M. K. (1991). The Big 5 personality dimensions and job-performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261–272.
- Barsade, S. G., Ramarajan, L., & Westen, D. (2009). Implicit affect in organizations. *Research in Organizational Behavior, 29*, 135–162.
- Bing, M. N., LeBreton, J. M., Davison, H. K., Migetz, D. Z., & James, L. R. (2007a). Integrating implicit and explicit social cognitions for enhanced personality assessment: A general framework for choosing measurement and statistical methods. *Organizational Research Methods, 10*, 346–389.
- Bing, M. N., Stewart, S. M., Davison, H. K., Green, P. D., McIntyre, M. D., & James, L. R. (2007b). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology, 92*, 722–744.
- Brief, A. P., & Weiss, H. M. (2002). Organizational behavior: Affect in the workplace. *Annual Review of Psychology, 53*, 279–307.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Christian, M. S., Garza, A. S., & Slaughter, J. S. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology, 64*, 89–136.
- Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*, 64–73.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–56.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago, IL: Rand McNally.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
- Davies, M., & Gardner, D. (2010). *A frequency dictionary of contemporary American English: Word sketches, collocates and thematic lists*. London: Routledge.
- Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology, 42*, 25–36.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.
- Edwards, J. R. (2008). To prosper, organizational psychology should ... overcome methodological barriers to progress. *Journal of Organizational Behavior, 29*, 469–491.
- Elliot, A. J. (2006). The hierarchical model of approach–avoidance motivation. *Motivation and Emotion, 30*, 111–116.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–327.
- Feldman, J. M., & Lynch, J. G., Jr. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73*, 421–435.
- Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Implicit attitudes spring from object-centered contexts. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 216–246). New York: Guilford Press.
- Fisher, C. D., & Ashkanasy, N. M. (2000). The emerging role of emotions in work life: An introduction. *Journal of Organizational Behavior, 21*, 123–129.
- Gatewood, R. D., & Feild, H. S. (2001). *Human resource selection* (5th ed.). Mason, OH: South-Western.
- Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology–Psychologie Canadienne, 50*, 141–150.
- Gawronski, B., Lebel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science, 2*, 181–193.

- George, J. M. (1991). State or trait: Effects of positive mood on prosocial behaviors at work. *Journal of Applied Psychology*, 76, 299–307.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypical beliefs. *Journal of Personality and Social Psychology*, 60, 509–517.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103, 265–275.
- Gur, R. C., & Sackeim, H. A. (1979). Self-deception: Concept in search of a phenomenon. *Journal of Personality and Social Psychology*, 37, 147–169.
- Haines, E. L., & Sumner, K. E. (2006). Implicit measurement of attitudes, stereotypes, and self-concepts in organizations: Teaching old dogmas new tricks. *Organizational Research Methods*, 9, 536–553.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21, 967–988.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1, 104–121.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369–1385.
- Izen, A. M., & Baron, R. A. (1991). Positive affect as a factor in organizational-behavior. *Research in Organizational Behavior*, 13, 1–53.
- Ito, T. A., & Cacioppo, J. T. (2007). Attitudes as mental and neural states of readiness: Using physiological measures to study implicit attitudes. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 125–158). New York: Guilford Press.
- James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, 1, 131–163.
- Johnson, R. E., & Lord, R. G. (2010). Implicit effects of justice on self-identity. *Journal of Applied Psychology*, 95, 681–695.
- Johnson, R. E., & Saboe, K. N. (2011). Measuring implicit traits in organizational research: Development of an indirect measure of employee implicit self-concept. *Organizational Research Methods*, 14, 530–547.
- Johnson, R. E., & Steinman, L. (2009). The use of implicit measures for organizational research: An empirical example. *Canadian Journal of Behavioural Science*, 41, 202–212.
- Johnson, R. E., & Tan, J. A. (2009). Explicit reasons for examining the implicit motive system. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 103–105.
- Johnson, R. E., Tolentino, A. L., Rodopman, O. B., & Cho, E. (2010). We (sometimes) know not how we feel: Predicting job performance with an implicit measure of trait affectivity. *Personnel Psychology*, 63, 197–219.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127, 376–407.
- Kanfer, R. (2009). Work motivation: Identifying use-inspired research directions. *Industrial and Organizational Psychology*, 2, 77–93.
- Kaplan, S., Bradley, J. C., Luchman, J. N., & Haynes, D. (2009). On the role of positive and negative affectivity in job performance: A meta-analytic investigation. *Journal of Applied Psychology*, 94, 162–176.
- Kim, J., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it*. Beverly Hills, CA: Sage.
- Kuhn, T. S. (1963). The essential tension: Tradition and innovation in scientific research. In C. W. Taylor, & F. Barron (Eds.), *Scientific creativity: Its recognition and development*. New York: Wiley and Sons.
- Latham, G. P., & Locke, E. A. (2008). Employee motivation. In J. Barling, & C. L. Cooper (Eds.), *The SAGE handbook of organizational behavior*, 1 (pp. 318–333). Los Angeles: Sage.
- Latham, G. P., Stajkovic, A. D., & Locke, E. A. (2010). The relevance and viability of subconscious goals in the workplace. *Journal of Management*, 36, 234–255.
- Leavitt, K., Fong, C. T., & Greenwald, A. G. (2011). Asking about well-being gets you half an answer: Intra-individual processes of implicit and explicit job attitudes. *Journal of Organizational Behavior*, 32, 672–687.
- LeBreton, J. M., Hargis, M. B., Griepentrog, B., Oswald, F. L., & Ployhart, R. E. (2007). A multidimensional approach for evaluating variables in organizational research and practice. *Personnel Psychology*, 60, 475–498.
- Lemm, K. M., Lane, K. A., Sattler, D. N., Khan, S. R., & Nosek, B. A. (2008). Assessing implicit cognitions with a paper-format implicit association test. In T. G. Morrison, & M. A. Morrison (Eds.), *The psychology of modern prejudice*. Hauppauge, NY: Nova Science Publishers.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27–66.
- Long, J. S. (1983). *Confirmatory factor analysis*. Beverly Hills, CA: Sage.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, 96, 690–702.
- Meyer, J. P., Stanley, D. J., Herscovitch, L., & Topolnysky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior*, 61, 20–52.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683–729.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36, 402–407.
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134, 565–584.
- Nosek, B. A., & Smyth, F. L. (2007). A multitrait–multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, 54, 14–29.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw Hill.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903.
- Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude–test validity in selection: A red herring. *Journal of Applied Psychology*, 66, 166–185.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait–multimethod matrices. *Applied Psychological Measurement*, 10, 1–22.
- Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, 77, 885–904.
- Smith, E. R., & Conrey, F. R. (2007). Mental representations are states not things: Implications for implicit and explicit measurement. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 247–264). New York: Guilford Press.
- Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108–131.
- Son Hing, L. S., Li, W., & Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial responses among aversive racists. *Journal of Experimental Psychology*, 38, 71–78.
- Staw, B. M., & Barsade, S. G. (1993). Affect and managerial performance: A test of the sadder-but-wiser vs. happier-and-smarter hypotheses. *Administrative Science Quarterly*, 38, 304–331.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test–performance of African-Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.
- Stokes, J. P., & Levin, I. M. (1990). The development and validation of a measure of negative affectivity. *Journal of Social Behavior & Personality*, 5, 173–186.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703–742.
- Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word–fragment completion are independent of recognition memory. *Journal of Experimental Psychology—Learning Memory and Cognition*, 8, 336–342.

- Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., & Johnson, R. E. (2012). Getting explicit about the implicit: A taxonomy of implicit measures and guide for their use in organizational research. *Organizational Research Methods, 15*, 553–601.
- Vargas, P. T., Sekaquaptewa, D., & von Hippel, W. (2007). Armed only with paper and pencil: “Low tech” measures of implicit attitudes. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 103–124). New York: Guilford Press.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.
- Way, J., & Johnson, R. E. (2010). Assessing employees’ regulatory focus using implicit measurement techniques. *Poster presented at the 25th Annual Society for Industrial and Organizational Psychology Conference, Atlanta, Georgia.*
- Weiss, H. M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior, 18*, 1–74.
- Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C., & Duncan, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review, 105*, 230–250.
- Wittenbrink, B., & Schwarz, N. (Eds.). (2007). *Implicit measures of attitudes*. New York: Guilford Press.