Measuring the Power of Learning.™

Research Report
ETS RR–18-24

Development of the Behaviorally Anchored Rating Scales for the Skills Demonstration and Progression Guide

David M. Klieger

Harrison J. Kell

Samuel Rikoon

Kri N. Burkander

Jennifer L. Bochenek

Jane R. Shore

December 2018

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Development of the Behaviorally Anchored Rating Scales for the Skills Demonstration and Progression Guide

David M. Klieger, Harrison J. Kell, Samuel Rikoon, Kri N. Burkander, Jennifer L. Bochenek, & Jane R. Shore

Educational Testing Service, Princeton, NJ

In this research, we developed behaviorally anchored rating scales (BARS) to evaluate the job performance of Zone Three jobs requiring middle-level skills and prior education ranging from vocational training to an associate's degree as well as work-related skills or experience. We ultimately identified 7 relevant job performance dimensions mainly based on prior research literature. The dimensions are thought to reflect socioemotional constructs (soft skills, "21st-century skills") considered vital to success today in Zone Three jobs. Managers of Zone Three employees helped us develop the final behavioral statements to anchor the 6 BARS points by generating approximately 430 critical incidents on which we based the initial behavioral statements that we wrote. Another group of managers confirmed the relevancy of the statements to the Zone Three workers they supervised, retranslated statements back into dimensions to confirm the dimensions to which they belong, and provided ratings of the effectiveness level that each statement represented. We modified statements to fill any gaps in effectiveness level coverage. Then, we asked leading experts in assessment to confirm the adequacy of the final statements, and after they did so, we created the final anchored scales. Both past literature and our own analyses indicate that these BARS generalize across economic sectors and thus potentially can provide substantial value to organizations that wish to assess, in an efficient and cost-effective manner, the performance of middle-skills employees with many different types of possible job descriptions.

Keywords BARS; Behaviorally Anchored Rating Scale; BSS; Behavioral Summary Scale; GRS; graphical rating scale; Zone Three; O*NET; training; job performance; noncognitive; psychosocial; socioemotional; 21st century skills; critical incident

doi:10.1002/ets2.12210

The goal of this research was to develop behaviorally anchored rating scales (BARS) to assess performance in occupations that the US Department of Labor/Employment and Training Administration's Occupational Information Network (O*NET) classifies as belonging to Job Zone Three (see O*NET Online, n.d.). Zone Three jobs require prior education ranging from vocational training to an associate's degree as well as work-related skills or experience [O*NET, 2017 (see O*NET Online, n.d.)]. Examples of Zone Three job roles include acute care nurse, electrician, loan officer, travel agent, and police detective. As discussed herein, this variety in the nature of Zone Three jobs does not preclude the use of the same measurement scales across them.

This development work was motivated by the US Congress's passage in 2014 of the Workforce Innovation and Opportunity Act (WIOA), which reauthorized the Workforce Investment Act. WIOA highlighted the fact that a broad array of skills, including noncognitive and personal workforce behaviors, knowledge, and competencies, are needed for performance and workplace success in jobs requiring at least a high school diploma but less than a 4-year college degree (Casner-Lotto & Barrington, 2006; National Research Council, 2012; U.S. Department of Labor, 2014).

This legislative concern reflects employers' intensified focus on 21st-century skills (which include constructs often classified as noncognitive, psychosocial, or socioemotional). Fueled in part by tremendous technological advances and increasing globalization, the US economy has undergone dramatic changes over the last five decades, shifting from a manufacturing-oriented economy to a more service- and information-based economy (Carnevale, Gainer, & Meltzer, 1988; Kemeny & Rigby, 2012). Many manufacturing jobs have been moved to other countries having lower cost labor, and the nature of most jobs—manufacturing and otherwise—has changed due to rapid transformations in technology and increased competition. As a result, the workforce skills that employers desire have changed. They seek workers with greater fluency in proficiencies like problem solving, teamwork, innovation, initiative, and communication. Failing to

Corresponding author: D. M. Klieger, E-mail: dklieger@ets.org

adequately support the development of these skills may compromise the nation's ability to remain competitive in the global marketplace (Burrowes, Young, Restuccia, Fuller, & Raman, 2014; Carnevale et al., 1988).

A key element of WIOA is the requirement for strategically aligning training with the needs of a prepared workforce that fits regional employer needs. WIOA also promotes a transparent evaluation system of such training that is "evidence based and data-driven" (U.S. Department of Labor, 2014). WIOA explicitly delineates that such evaluations be conducted not just of job acquisition and retention but also of credentials and specific measurement of skills developed in training programs that are relevant to in-demand jobs. Therefore the more immediately intended application of the BARS is to aid training and development programs in assessing the success of training given to individuals who subsequently perform Zone Three jobs; however, the BARS were also designed to be usable across contexts in which a main goal is to evaluate the job performance of workers in Zone Three jobs. For example, these BARS potentially could be used as a system for end-of-year performance evaluations.

Zone Three roles exist across multiple job sectors (e.g., administrative, health care, human services, manufacturing/technology/construction, and agricultural). While Zone Three jobs may include management positions, the BARS reflect the reality that most Zone Three jobs lack a significant formal management component. A Zone Three job can have more than one hierarchical level (e.g., both a more junior-level and a more senior-level position), provided that those levels do not require substantially different amounts or types of education, training, experience, or job duties.

The BARS were designed primarily to assess skills that prior research has demonstrated to be important to success in performing many types and levels of jobs. These skills constitute the behavioral expression of personality, motivational, attitudinal, and self-regulatory constructs thought to be critical determinants of job performance (Barrick & Mount, 1991; Harrison, Newman, & Roth, 2006; Judge & Bono, 2001; Judge & Ilies, 2002). These are not the kinds of skills or abilities that traditional cognitive ability assessments (e.g., the Wonderlic, the Raven's Progressive Matrices) are designed primarily to measure; that is, they are not skills or abilities such as general mental ability (i.e., IQ); mental processing speed; working memory; mathematical, verbal, writing, critical reasoning, mechanical, or spatial ability; or declarative or procedural knowledge of a subject area like biology, the components of a mortgage loan, or the procedures for making a lawful arrest. Nor are they abilities or skills manifested physically, such as strength, speed, or coordination.

Sometimes, the skills that the BARS are designed to measure are referred to as psychosocial, socioemotional, noncognitive, or personality based in nature (Kyllonen, 2016; Robbins et al., 2004). Yet, it is impossible for any valid and useful assessment in the real world to avoid measuring traditional cognitive and physical attributes to at least some extent. Generally, the behavioral expression of noncognitive constructs that an assessment intends to target is inextricably intertwined with the possession of certain factual knowledge, cognitive skills, and physical capacities. Extending effective help to a coworker in completing a task at work requires knowing how to complete the task. Reading and responding effectively to a manager's subtle facial gestures requires adequate mental processing speed. In Zone Three positions that generally require physical activity (e.g., police detective), demonstration of initiative may involve salient physical behaviors, such as making extra physical efforts. Nevertheless, the intention of the BARS is to assess personality, motivational, attitudinal, and self-regulatory constructs thought to be critical determinants of job performance.

Job Performance

Because training is "a planned intervention that is designed to enhance the determinants of individual job performance" (Campbell & Kuncel, 2011, p. 278), the history of training performance is a part of the history of job performance, and vice versa. The scientific measurement of job performance (and thus training performance) began its maturation during World War I, when Walter Dill Scott developed a rating scale to hire for sales positions (Farr & Levy, 2007). Curiously, systematic attention to the structure and content of job performance lagged behind the actual measurement of job performance for many decades (Austin & Villanova, 1992; Motowidlo & Kell, 2013). While serious efforts were made to better understand the nature of job performance in the 1970s (Arvey & Mussio, 1973), it was not until the late 1980s and the 1990s that the structure and content of job performance were methodically and comprehensively studied in ways that have gained wider acceptance (Borman & Motowidlo, 1993; Campbell, 1990).

Campbell's (1990) multifactor model was one of the first attempts to identify performance domains that could be generalized across job types. The eight factors are job-specific proficiency, non-job-specific task proficiency, written and oral communication, demonstrating effort, maintaining personal discipline, facilitating team and peer performance, supervision, and management and administration. Borman and Motowidlo (1993) sought to make a distinction between task

performance and contextual performance, the former being related directly to performance of explicit job requirements, the latter to other behaviors that promote organizational effectiveness. Coleman and Borman (2000) built on this model, further identifying three categories of contextual performance: interpersonal support, organizational support, and job-task conscientiousness. These three categories were further refined by Borman, Penner, Allen, and Motowidlo (2001) into personal support, organizational support, and conscientious initiative. Other behavioral dimensions that have received significant scholarly attention are organizational citizenship behavior (LePine, Erez, & Johnson, 2002; Organ, 1988, 1997; Podsakoff, MacKenzie, Moorman, & Fetter, 1990), counterproductive behavior (Robinson & Bennett, 1995; Robinson & O'Leary-Kelly, 1998; Sackett, 2002), and adaptive performance (Allworth & Hesketh, 1999; Griffin & Hesketh, 2003; Pulakos, Arad, Donovan, & Plamondon, 2000).

History of Ratings-Based Job Performance Scales

Although the performance of work predates recorded human history, the best method for evaluating job performance is not well established (Austin & Villanova, 1992). The first documented use of rating scales for evaluating psychological variables is believed to be the work of Christian Thomasius, a philosopher who had devised his own theory of personality some 300 years ago (McReynolds & Ludwig, 1987). Rating scales came into wide use by psychologists through their application by phrenologists in the mid-19th century (Bakan, 1966) and by Francis Galton in the late 19th century (Guilford, 1954). Supervisors evaluating their subordinates' performance using some type of rating scale has become the most frequently used method of appraising job performance (Borman, 1991).

Graphical rating scales (GRS) are the most common type of (Cascio & Aguinis, 2004; Guion, 2011), and also perhaps the longest serving (Freyd, 1923; Miner, 1917; Paterson, 1922, 1923), ratings-based performance appraisal devices. GRS do not have a wholly standardized format but, fundamentally, consist of a list of job performance dimensions, job-relevant human characteristics (e.g., cooperation, flexibility, initiative, sociability; Borman & Dunnette, 1975; Freyd, 1923), or both, accompanied by an evaluative continuum (e.g., below average to outstanding, very high to very low) upon which supervisors are asked to indicate their judgments about target employees. Comprehensive GRS can be found in Guion (2011, p. 451, Figure 13.1).

Despite the long standing and wide ranging use of GRS, they have been criticized for being vague or ambiguous (Latham, Fay, & Saari, 1979; Schwab, Heneman, & DeCotiis, 1975) and, as a consequence, contributing construct-irrelevant variance (Messick, 1989) to scores derived from them. This invalid variance occurs as a consequence of raters' idiosyncratic interpretations of the evaluative dimensions and the continuum levels upon which those dimensions are being appraised (Smith & Kendall, 1963). For example, different supervisors may have differing interpretations of what "takes initiative" constitutes, leading them to rate the same employee differently, even if they have observed identical samples of that employee's behavior. Additional construct-irrelevant variance may be contributed if supervisors have differing conceptions of what it means to perform "above average" or "below expectations," and for each dimension. Even when these verbal labels for differing levels of effectiveness are replaced with numbers, the possibility of differential interpretation by raters remains (Borman, 1986).

Behaviorally Anchored Rating Scales

BARS were invented to address the deficiencies of prior attempts at GRS. The key feature of BARS is that they provide concrete behavioral examples of different levels of performance. These examples serve as both explicit standards that raters can use when evaluating an employee's performance and implicit definitions of what performance comprises at different levels of effectiveness. Rather than it being left up to the rater to interpret what, for example, "above average" performance means, BARS *show* the rater via a behavioral exemplar. A testament to the utility of BARS is their application outside the job performance appraisal context for which they were originally developed. A nonexhaustive list of the constructs BARS that have been developed over the past 50 years includes employment interview performance (Levashina, Hartwell, Morgeson, & Campion, 2014), the Big Five personality traits (Muck, Hell, & Höft, 2008), motivation (Landy & Guion, 1970), classroom teamwork (Ohland et al., 2012), and evaluating teaching practice (Martin-Raugh, Tannenbaum, Tocci, & Reese, 2016). Given that BARS are often intrinsically tied to specific jobs and organizations, they have also been suggested to serve as a foundation for feedback and training programs (Blood, 1974; Campbell, Dunnette, Arvey, & Hellervik, 1973; Hom, DeNisi, Kinicki, & Bannister, 1982).

General Development Procedure

A variety of procedures exist to develop BARS, but the great majority of these approaches hew closely to the original technique introduced by Smith and Kendall (1963) (for more comprehensive reviews, see Kell et al., 2017; Lee & Tindal, 1996). Typically, the sequence has seven steps:

1. The process begins with the application of the critical incident technique (Flanagan, 1954), in which subject matter experts (SMEs; knowledgeable job incumbents) provide examples of workplace behavior. Critical incidents typically depict specific, concrete, highly effective, and highly ineffective workplace behaviors.
2. Inductive creation of job performance categories. The BARS developers edit the incidents into a common format and eliminate redundancies, then form groups of incidents based on their content similarities. The developers create labels and definitions for these groups according to these content similarities. These inductively derived categories form a preliminary job performance taxonomy that is inherently, defined in terms of workers' actions.
3. Often called *retranslation*, a second group of SMEs (not overlapping with those that generated the critical incidents) is given the critical incidents in randomized order, along with the list of performance categories and their definitions. These SMEs place each incident into the performance category in which they believe it best fits.
4. BARS developers compute agreement statistics for each critical incident for their placement into the performance domains; incidents that do not meet some predetermined agreement standard are discarded.
5. Surviving incidents are given to a third nonoverlapping group of SMEs. These SMEs rate the incidents for effectiveness.
6. BARS developers compute the mean effectiveness values for each incident and use the standard deviation of the SMEs' ratings to index the degree of agreement. Incidents that do not meet some predetermined agreement standard are discarded.
7. Remaining incidents are used to prepare the final BARS, with the mean effectiveness ratings of the incidents determining their placement on the effectiveness continuum for the performance category to which they have been assigned. These incidents *are* the "anchors" in BARS.

Benefits

The major advantage attributed to BARS is based on the rational consideration that they reduce construct-irrelevant variance in performance appraisal ratings through their emphasis on specific, concrete, observable behaviors as a means of defining the dimensions to be judged and anchoring the evaluative continuum (Smith & Kendall, 1963); these examples serve as common reference points on which raters base their judgments (Jacobs, Kafry, & Zedeck, 1980). Consistent with this reasoning, multiple studies have demonstrated that ratings using BARS sometimes (but not always) exhibit less measurement bias (e.g., halo, leniency, range restriction) than those produced using other types of scales (Bernardin, LaShells, Smith, & Alvares, 1976; Borman & Dunnette, 1975; Campbell et al., 1973; Keaveny & McGann, 1975; Rumsey & Mietus, 1983; Tziner, 1984). The psychometric superiority of BARS, when demonstrated, is often attributed to the rigor with which they are developed (Bernardin, 1977; Borman, 1986, 1991; Landy & Farr, 1980). In addition to the psychometric benefits BARS sometimes demonstrate, there is also evidence for behavior-based ratings scales being linked to more favorable attitudinal reactions, including perceptions of fairness and justice (Ivancevich, 1980; Roch, Sternburgh, & Caputo, 2007; Tziner & Kopelman, 2002), and for being effective when used as a basis for planning behavioral change (Hom et al., 1982). Furthermore, although the evidence comes from the job interview rather than the job performance context, there are indications that use of BARS may decrease bias against protected groups (Reilly, Bocketti, Maser, & Wennet, 2006).

The utility of BARS is also due to the fact that their components are ultimately traceable to SMEs, who are, by definition, intimately familiar with the target job (Campbell et al., 1973). Indeed, it has been argued that because the source of the elements composing BARS is job experts' judgments, they automatically possess job relevance—a critical component of legal defensibility in adverse impact cases (Jeanneret & Zedeck, 2010; Landy, Gutman, & Outtz, 2010). For example, Jacobs et al. (1980) noted,

> Perhaps the strongest attribute of the BARS methodology is its ability to yield job analysis information performed by the people who know the job best and written in their language. By generating and evaluating behavioral items

> necessary for the final format, the BARS methodology results in explicit statements regarding requisite job behaviors and their perceived value. On this level BARS item generation can be seen as meeting the criterion of relevancy (p. 606).

Other aspects of BARS lend themselves to legal defensibility concerns as well (Nathan & Cascio, 1986; Pulakos, 2007). First, SMEs' provision of critical incidents often serves as a job analysis, as the content of the critical incidents constitutes a description of the content of the job itself. Second, BARS inherently feature behavior-based dimensions rather than personality traits, which some courts may find to be too subjective (*Amini v. City of Minneapolis*, 2011; *Pierce v. Board of Regents of the University System of Georgia*, 2011). Third, BARS emphasize specific job performance dimensions rather than global or "overall" evaluations of workplace effectiveness.

Behavioral Summary Scales

The very specific behavioral anchors that are the central feature of BARS sometimes prove difficult for raters because of their extreme specificity (Atkin & Conlon, 1978). Behavioral summary scales (Borman, 1979, 1986; Borman, Hough, & Dunnette, 1976) are a subtype of BARS created to address this issue. Although not fundamentally different from traditional BARS, in BSS development procedures, the critical incidents surviving after the sixth step described earlier are further content analyzed. Developers identify themes common to multiple incidents that have been sorted into the same performance dimension and are of approximately equal effectiveness; brief behavioral statements that summarize these themes are written. These statements are more general than the single critical incidents that form the anchors in traditional BARS but are still grounded in workplace actions and are not as abstract as the adjectives frequently used to assess personality traits. Additionally, the behavioral statements often anchor multiple scale points. The rationale for this procedural modification is that even when there is high agreement on an incident's effectiveness, there is still likely enough variability in SMEs' judgments that reasonable raters may disagree on whether a behavior is "truly" indicative of, for example, performance at an effectiveness level of 6 or an effectiveness level of 7 (Bernardin & Smith, 1981). Using summary statements as anchors for multiple scale points is intended to account for the possibility of these flip-flops. The BARS that we developed are of the BSS type in order to benefit from the aforementioned advantages.

Generalizability

BARS are designed generally for evaluating the performance of a specific job or job within a job family. In this project, we innovate that tradition by developing BARS for an entire job zone. (We were able to identify one study, Goodale & Burke, 1975, that developed BARS for a range of jobs at a hospital.) As discussed, Zone Three cuts across sectors and covers a wide variety of jobs. While the summary descriptions of job tasks, work activities, work context, and other details vary across Zone Three jobs, there are behaviors across Zone Three jobs that universally (or nearly universally) will contribute to or detract from successful job performance. The dimensions into which these behaviors can be organized will generalize widely. As Campbell (2012) elaborated,

> At a particular level of specificity/generality, research has shown that particular sets of actions (e.g., refraining from substance abuse, showing consideration for coworkers, setting goals with subordinates) contribute to goal accomplishment in virtually any organization … At a particular level of specificity, there is a virtual consensus about the latent variables that comprise individual performance at work. Going only a small step further, it is strongly suggested that this latent structure is invariant across work roles, organizational levels, organizational structures, organizational contexts, and so on, and so on. This is not an argument that the importance or utility of individual differences on each latent variable is the same across work roles, organizational levels, and situations. For example, not all jobs would have a significant management component (although this number may be larger than we think), and communication, as specified there, might be a critical component of only a small percentage of jobs.
>
> The assertion of invariance is also not an argument that individuals won't adapt their performance behavior to changing contexts or situations. It is, rather, that such adaptations or differential emphases across situations are best described within the consensus latent structure framework (p. 4).

Not only is the taxonomic structure of job performance generalizable but to a large extent so too is the effectiveness of behaviors across a broad array of jobs. As Campbell (2012) stated,

> Also, asserting that the latent structure of performance is invariant across levels of work roles is not synonymous with saying that the actions comprising high and low performance on the dimension are invariant. However, it comes close. When is technical performance expertise not good? When is expert communication not good? When is a lack of CWB [counterproductive workplace behaviors] not good? When are extra effort and initiative not good? When is it not good to be highly competent on the components of leadership and management, even if the relative utility of the subfactors varies across situations? The only possible exceptions are with regard to the influence of culture. For some of the leadership subfactors, the same actions may have different effects on peers or subordinates as a function of cultural values. However, even here, the numbers of such cultural interactions may be relatively small (p. 31).

The BARS dimensions and the behavioral anchors that we identified and developed are consistent with J. P. Campbell's perspective that the structure of job performance and effectiveness of work behaviors are invariant across different jobs. Behavioral expression of the dimensions universally (or nearly universally) contributes to or detracts from successful job performance. It is important to recognize this scientific reality and to leverage its utility for organizations that do not need the cost and complexity of numerous taxonomies and sets of scales to assess the performance of employees across different jobs. As discussed later, we analyzed the extent to which managers across different job sectors agree on dimension relevance, statements' dimensional assignments, and the effectiveness represented by the statements.

Method and Analyses

Creation of a Job Performance Taxonomy

We conducted a thorough review of the literature regarding job performance taxonomies and originally identified eight common domains of workplace performance applicable for middle-skill workers (see Appendix A, which extends the dimensional mapping approach of Koopmans et al., 2011). After further review, we consolidated leadership and management into one domain, because we found it challenging to imagine a sufficient number of realistic and distinct examples of both leadership and management behaviors for Zone Three workers. The deliberation process yielded seven performance domains, for which we created an operational definition to guide our work.

Service

Service (Arvey & Mussio, 1973; Bartram, 2005; Bartram, Robertson, & Callinan, 2002; Chan, 2006; Golubovich, Su, & Robbins, 2017; Kurz & Bartram, 2002; Maxham III, Netemeyer, & Lichtenstein, 2008; Shore, Lentini, Rikoon, Seybert, & Noeth, 2016; Tett, Guterman, Bleier, & Murphy, 2000; Viswesvaran & Ones, 2000) was defined as.

> conducting oneself in a polite, patient, cooperative manner with individuals external to the organization (e.g., clients, customers, inspectors/auditors, vendors); acting to meet these individuals' needs; following through with these individuals to get the job done well; appropriately managing these individuals if they become difficult.

It was noted that the service domain may not apply to all Zone Three jobs, so this domain is denoted as "if applicable."

Initiative and Work Ethic

Initiative and work ethic (Allworth & Hesketh, 1999; Arvey & Mussio, 1973; Bakker, Demerouti, & Verbeke, 2004; Bartram, 2005; Bartram et al., 2002; Borman & Brush, 1993; Borman & Motowidlo, 1993; Burrus, Jackson, Xi, & Steinberg, 2013; Campbell, 1990, 2012; Campbell et al., 1990; Campbell, McHenry, & Wise, 1990; Casner-Lotto & Barrington, 2006; Engelbrecht & Fischer, 1995; Fluegge, 2008; Golubovich et al., 2017; Greenslade & Jimmieson, 2007; Griffin, Neal, & Parker, 2007; Hedge, Borman, Bruskiewicz, & Bourne, 2004; Hunt, 1996; Kurz & Bartram, 2002; Lance, Teachout, & Donnelly, 1992; Luo, Shi, Li, & Miao, 2008; Mael et al., 2010; Maxham III et al., 2008; Michel, 2000; Murphy, 1989; Pulakos et al.,

2002; Renn & Fedor, 2001; Rollins & Fruge, 1992; Rotundo & Sackett, 2002; Shore et al., 2016; Sinclair & Tucker, 2006; Tett et al., 2000; Van Dyne, Jehn, & Cummings, 2002; Viswesvaran, 1993; Warner, Gates, Christeson, & Kiernan, 2011; Wisecarver, Carpenter, & Kilcullen, 2007) is a broad category that includes behaviors that might otherwise be considered general task performance. We defined initiative and work ethic as "anticipating problems and solving them before they arise; persisting in difficult, long-term, or unpleasant tasks until they are successfully accomplished; completing tasks effectively (e.g., accurately, efficiently, in a timely manner)."

Communication Skills

Communication skills (Bartram, 2005; Bartram et al., 2002; Borman & Brush, 1993; Campbell, 1990, 2012; Casner-Lotto & Barrington, 2006; Chan, 2006; Hedge et al., 2004; Kurz & Bartram, 2002; Rollins & Fruge, 1992; Tett et al., 2000; Viswesvaran, 1993; Warner et al., 2011) are conceived to include both oral and written communication and are defined as "speaking, writing, and listening effectively; understanding and appropriately responding to people's actions in face-to-face communications, by email, teleconferences, and/or video conferences; reading others' body language."

Flexibility and Resilience

Flexibility and resilience (Allworth & Hesketh, 1999; Bartram, 2005; Bartram et al., 2002; Burrus et al., 2013; Griffin et al., 2007; Hunt, 1996; Kurz & Bartram, 2002; Pulakos et al., 2002; Rollins & Fruge, 1992; Shore et al., 2016; Sinclair & Tucker, 2006; Tett et al., 2000; Warner et al., 2011) captures behaviors that might otherwise be described as adaptive performance. It is defined as "adapting well to unclear or changing work demands; handling stress appropriately; accepting criticism or feedback without getting discouraged or overly defensive; remaining focused on getting the job done when faced with challenges and setbacks."

Problem-Solving Skills

Problem-solving skills (Bartram, 2005; Bartram et al., 2002; Casner-Lotto & Barrington, 2006; Chan, 2006; Engelbrecht & Fischer, 1995; Fluegge, 2008; Jiambalvo, 1979; Kurz & Bartram, 2002; Shore et al., 2016; Van Dyne et al., 2002; Warner et al., 2011) are defined as "using information, knowledge, and reasoning to solve problems; thinking critically and creatively; using good judgment when making decisions; looking for new information, when necessary, to solve problems."

Responsibility

Responsibility (Allen, 2008; Arvey & Mussio, 1973; Bartram, 2005; Bartram et al., 2002; Borman & Brush, 1993; Burrus et al., 2013; Burton, Pransky, Conti, Chen, & Edington, 2004; Campbell, 1990, 2012; Campbell, McHenry, & Wise, 1990; Casner-Lotto & Barrington, 2006; Engelbrecht & Fischer, 1995; Escorpizo, 2008; Golubovich et al., 2017; Hedge et al., 2004; Hunt, 1996; Kurz & Bartram, 2002; Luo et al., 2008; Murphy, 1989; Rollins & Fruge, 1992; Rotundo & Sackett, 2002; Shore et al., 2016; Sinclair & Tucker, 2006; Tett et al., 2000; Viswesvaran, 1993; Viswesvaran & Ones, 2000; Warner et al., 2011; Wisecarver et al., 2007) captures behaviors that might otherwise be described as the opposite of counterproductive behaviors. We defined this domain as "being accountable for one's own duties and actions; following safety and other rules, procedures, and policies; maintaining high standards of personal conduct and professionalism (e.g., being ethical, respectful)."

Teamwork and Citizenship

Teamwork and citizenship (Arvey & Mussio, 1973; Bartram, 2005; Bartram et al., 2002; Borman & Brush, 1993; Burrus et al., 2013; Campbell, 1990, 2012; Casner-Lotto & Barrington, 2006; Chan, 2006; Engelbrecht & Fischer, 1995; Golubovich et al., 2017; Jiambalvo, 1979; Kurz & Bartram, 2002; Lance et al., 1992; Luo et al., 2008; Michel, 2000; Murphy, 1989; Rollins & Fruge, 1992; Shore et al., 2016; Tett et al., 2000; Viswesvaran, 1993; Warner et al., 2011; Wisecarver et al., 2007) includes behaviors that might also be considered organizational citizenship behavior. This domain is defined as "working well with all members of the organization, both individually and in groups; demonstrating respect for different opinions, customs,

and preferences; actively participating in formal and informal group processes; being cooperative, helpful, and supportive to others."

Leadership and Management

Leadership and management (Arvey & Mussio, 1973; Bartram, 2005; Bartram et al., 2002; Borman & Brush, 1993; Campbell, 1990, 2012; Campbell, McHenry, & Wise, 1990; Casner-Lotto & Barrington, 2006; Engelbrecht & Fischer, 1995; Hedge et al., 2004; Jiambalvo, 1979; Kurz & Bartram, 2002; Luo et al., 2008; Mael et al., 2010; Rollins & Fruge, 1992; Tett et al., 2000; Viswesvaran, 1993; Wisecarver et al., 2007) deals with the guidance and support of others, originally conceived as behavior that could be observed in the context of a formal supervisory relationship or informally with peers. We defined this as "leading, directing, mentoring, evaluating, giving feedback, and setting work-related goals and expectations for others (note: does not have to be an official requirement of one's job)." However, subsequent feedback from participants at a workshop to develop critical incidents based on this dimension (see later) indicated that managers of Zone Three employees did not commonly or consistently identify relevant behaviors for this dimension, so it was removed from further analyses.

Development of Critical Incidents

After deciding upon a job performance taxonomy, we then proceeded to use the critical incident technique (Flanagan, 1954) to begin the process of generating behavioral statements to represent various levels of effectiveness in performance with respect to each of the dimensions. To facilitate recruitment of initial SMEs, or managers and supervisors of Zone Three employees — managers and supervisors who would participate in the BARS development — the development team created a one-page document outlining details and benefits of participation in the BARS Critical Incident Development Workshop. Members of the development team met with local contacts, providing the one-page document and the targeted industries set forth in a sampling plan. The sampling plan (described later) was developed through a systematic review of key Zone Three industries, identifying specific economic subsectors that were most representative of Zone Three jobs. Recruitment began with several workforce training programs in a metropolitan area of a city in the northeastern United States. Those targeted for recruitment were employers who engage in some way with such workforce training programs. As these programs had established relationships through internship and job opportunity programs, they aided by sharing their networks to support the recruitment process. Recruitment then involved reaching out to local business, trade, and training organizations to aid in identifying managers and supervisors of Zone Three workers. This included local Workforce Investment Boards, a trade union for health care, and an IT company whose mission was to aid in finding jobs for those trained in IT. The city government in which the workshop took place also aided in recruitment of its own employees to ensure broad representation across target industries.

A total of 24 managers of Zone Three employees were recruited and participated in the Critical Incident Development Workshop to support the eventual development of the BARS. These individuals came from administration, health care, human services, and manufacturing/technology/construction. These industries were targeted as they represent the categories that include the largest percentage of jobs that make up Zone Three (U.S. Department of Labor, 2015). Of the 24 participants, 5 (21%) were male. The majority of participants ($n = 16$) fell into the age range between 25 and 44 years. Five (21%) were White, 14 (58%) were African American, 2 (8%) were Hispanic, 2 (8%) were American Indian, and 3 (13%) did not provide a defined racial group. The group was varied in their educational background as well: four (17%) had a high school credential, five (21%) had a professional certificate, four (17%) had an associate's degree, five (21%) had a bachelor's degree, and six (25%) had a graduate degree.

Managers participating in the BARS development were asked to join a Critical Incident Workshop. Participants were instructed to provide critical incidents that are examples of highly ineffective, just good enough, and highly effective behavior in the performance categories defined. Furthermore, they were to provide (a) details about the specific situation in which the behavior happened, (b) details about the person's specific behavior itself, and (c) details about the specific outcome of that person's behavior. Following is an example of a critical incident produced for the responsibility dimension:

> An employee transitions to a new department. Prior to the transition, staff takes it upon self to reach out to new department to inquire about its strategies/tips. The employee is able to smoothly transition into the new position.

A total of approximately 430 critical incidents were collected from the participating managers.

Creation of Behavioral Statements

The next step in the BARS methodology was to distill the incidents into exemplars of work behavior (behavioral statements). We followed established procedures to develop behavioral statements (Borman, 1979; Borman et al., 1976; Hedge et al., 2004); several supplemental steps were incorporated to lend the process additional rigor. In multiple rounds, members of the research team revised statements; critical incidents were edited for clarity and redundancy; common elements of multiple incidents were abstracted and captured by writing more general behavioral statements summarizing their content; and random samples of the statements for each performance domain were reviewed by team members and further edited for clarity and to ensure that they were framed generally enough to apply to Zone Three jobs generally.

The specific steps were as follows:

1. The critical incidents generated by the workshop SMEs were randomly assigned to six members of the R&D team. They edited the incidents for coherence and relevance and eliminated incidents they judged to be redundant.
2. The edited critical incidents were randomly assigned to the same R&D staff members, although no team member received the same incidents that she or he had been responsible for editing. They examined the content of the edited incidents and grouped them according to themes shared across their behavioral aspects. Team members then wrote succinct statements intended to capture the essence of these common, overarching themes.
3. The behavioral statements, along with the edited critical incidents from which they were derived, were randomly assigned to the six members of the R&D team; team members did not receive the behavioral statements they were responsible for generating. Team members reviewed the statements in light of the incidents on which they were based and, when they judged it necessary, made suggestions for further refining the statements in light of their source material.
4. The team met as a whole to discuss the statements that reviewers in Step 3 found to be problematic. The final form of these statements was reached via group consensus.

The following is an example of an edited behavioral statement derived from the critical incident described previously:

> Prior to transitioning to new department, reaches out to relevant coworkers to inquire about strategies for new position.

The 398 statements derived from Step 4 were the stimuli that SMEs in Study 2 rated for effectiveness and retranslated into the performance categories.

Managers of Zone Three workers could not create nearly enough incidents for the leadership and management dimension to facilitate the creation of behavioral statements. On the basis of the participants' written workshop responses, we concluded that they could not recall enough examples in which leadership and management skills were salient to success as a Zone Three employee. Given that Zone Three employees were in positions junior to the participants, the participants might have mentally classified leadership and management behaviors as instead belonging to other dimensions, such as teamwork and citizenship or initiative and work ethic.

Retranslation, Relevance, and Effectiveness Ratings

To verify that each behavioral statement was a relatively pure indicator of a single performance dimension, a new group of SMEs sorted each of the behavioral statements back into the dimensions using an online survey of the behavioral statements. For the sake of efficiency, the same survey also asked new SMEs to rate the statements for relevance to the jobs of Zone Three employees that they managed as well as for the level of job performance effectiveness that they represented.

In return for a $150 online gift card, each participant consented and responded to the online survey. Each participant initially completed a screening survey to verify that his or her job role was appropriate for the study. Once this was confirmed, the participant was assigned to receive one of three forms of the survey. The survey was broken into three forms due to length and concerns about survey fatigue. Each form contained approximately 150 statements for individuals to classify into one of eight behavioral dimensions, rate for effectiveness level on a 6-point scale ranging from 1 (*highly*

Table 1 Examples of Participant Job Titles by Industry

Industry	Job titles
Administration	HR manager
	Managing creative director
	Help desk support
	Director of finance and administration
Health care	Clinic supervisor
	Population health specialist
	Personal care specialist
	Project manager (for a health care company)
Human services	GED teacher
	Detective sergeant
	Registrar
	Director of theatre
Manufacturing/technology/construction	Director of workforce collaboration technologies
	Information analyst
	Software engineer
	Director of facilities

ineffective) to 6 (*highly effective*), and indicate whether a given statement was relevant to the respondent as a manager of Zone Three workers.

To recruit for the BARS retranslation survey, the development team expanded its reach to national workforce organizations, state labor departments, community college consortia, national training programs, and trade unions. Through professional networks, representatives from these organizations were contacted and provided with details and benefits of participation and with target industries. These representatives assisted in recruiting a total of 71 managers of Zone Three employees. (For purposes of analysis, we ultimately obtained usable data from 65 of them.) These managers rated statements for relevance and effectiveness and sorted them into one of the performance dimensions in a continued online survey activity. These managers represented core economic sectors: administration, health care, human services, and manufacturing/technology/construction. Members of our research team discussed potential industrial classifications for each participant job title, grouping them into industry categories upon reaching consensus. Table 1 provides several examples of the job titles assigned to each industry (e.g., HR manager classified as administration, population health specialist classified as health care).

Table 2 shows the breakdown of participant raters by their respective industry classifications, displaying both demographic statistics and the distribution of experience in their current field. After efforts to diversify the background of the SME group, the sample was predominantly female (63%), White (91%), and between the ages of 35 and 44 years (49%), with 20% in administration, 15% in health care, 31% in human services, and 34% in manufacturing/technology/construction. Across all industries, 50% or more of the participants had been in the field for 10+ years.

The behavioral statements were clearly relevant to the participants overall and across economic sectors. By averaging across behavioral statements for each respondent, Table 3 shows the percentage of respondents who agreed that the behavioral statements for each dimension were in fact relevant for the Zone Three workers. The lowest total overall percentage agreement was 84% (for service), mainly because the percentage agreement for service in the administration sector was 74%. Yet, even these lowest values for agreement show that the statements are generally relevant overall and across the economic sectors in Zone Three jobs.

Two decision criteria dictated whether a statement would be retained for consideration as a potential component of the final scales. First, 50% or more of Study 2's SMEs overall had to have sorted the statement into the same performance category (cf. Schwab et al., 1975). Results were aggregated and computed by dimension before being analyzed by economic sector. The extent to which raters agreed that the behavioral statements eventually assigned to a given competency should in fact be aligned with that competency (hereafter referred to simply as "agreement") is displayed in Tables 4 – 10 (one for each competency). The highest level of agreement for the set of items within each behavioral competency scale was always with the given scale itself. This reflected the process we used to create the behavioral competency scales, in which only items demonstrating overall majority agreement (≥50%) were considered for inclusion.

Table 2 Participant Table by Industry

	Administration		Health care		Human services		Manufacturing/ technology/construction		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender										
Male	2	15	1	10	4	20	17	77	24	37
Female	11	85	9	90	16	80	5	23	41	63
Age (years)										
18–24	0	0	1	10	0	0	0	0	1	2
25–34	1	8	2	20	8	40	3	14	14	22
35–44	7	54	6	60	9	45	10	45	32	49
45–54	4	31	1	10	1	5	6	27	12	18
55–64	1	8	0	0	1	5	3	14	5	8
65+	0	0	0	0	1	5	0	0	1	2
Race										
White	12	92	9	90	18	90	20	91	59	91
African American	0	0	1	10	1	5	1	5	3	5
Hispanic	1	8	0	0	1	5	0	0	2	3
Multiracial	0	0	0	0	0	0	1	5	1	2
Industry duration (years)										
<5	3	23	2	20	2	10	2	9	9	14
5–10	3	23	3	30	6	30	4	18	16	25
10+	7	54	5	50	12	60	16	73	40	62
Total	13	100	10	100	20	100	22	100	65	100

Note. Percentages are by column.

Table 3 Relevance of Behavioral Statements to the Subject Matter Experts, by Sector/Industry and Overall

	Sector/industry (%)				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Overall
Communication skills	93	86	95	97	94
Flexibility and resilience	87	92	97	97	94
Initiative and work ethic	95	84	79	95	88
Problem-solving skills	100	100	92	100	98
Responsibility	91	85	96	95	93
Service	74	83	93	82	84
Teamwork and citizenship	84	83	92	90	88

There were a few scales in which some differences were exhibited in agreement statistics across sectors (e.g., problem-solving skills, Table 8; responsibility, Table 9). Examining problem-solving skills in more detail, although the greatest number of raters in administration agreed that these statements aligned with the problem-solving skills competency (44%), 31% thought the statements could also belong to the communication skills area (a discrepancy not observed in any of the other sectors). This could indicate that, among those in administration roles, there is a conceptual link between problem solving and communication skills that is less prominent in the other industries represented.

Considering initiative and work ethic (Table 5), communication skills (Table 6), flexibility and initiative (Table 7), and teamwork and citizenship (Table 10), the differences observed between industry subgroups were smaller, respectively, than those in the remaining three competency areas; that is, the proportions of each subgroup assigning each type of rating were similar enough to one another that, for the aforementioned constructs and dimensions, it appeared that each industry conceptualized the dimensions in approximately the same manner. For example, differences between the highest and lowest agreement statistics among industry subgroups across the four areas ranged from 6% (initiative and work ethic, Table 5) to 16% (flexibility and initiative, Table 7). In contrast, differences between the highest and lowest agreement statistics among industry subgroups across the remaining three areas ranged from 22% (service, Table 4) to 36% (both problem-solving skills and responsibility, Tables 8 and 9, respectively).

Table 4 Service

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
N_i	13	10	20	22	65
N_r	39	30	60	66	195
Communication skills (%)	8	7	0	3	4
Flexibility and resilience (%)	3	0	7	8	5
Initiative and work ethic (%)	8	7	8	12	9
Management/leadership (%)	0	0	3	0	1
Problem-solving skills (%)	5	3	10	3	6
Responsibility (%)	13	3	12	8	9
Service (%)	**59**	**80**	**58**	**65**	**64**
Teamwork and citizenship (%)	5	0	2	2	2

Note. N_i is the number of unique individuals who provided responses to the items chosen for the scale, whereas N_r reflects the number of behavioral statements rated by each individual. No item was answered more than once by the same individual. Percentages show the overall percentages of statements that were classified within each BARS dimension across all respondents (i.e., rates of respondent agreement that the pool of behavioral statements in question was aligned with each dimension). Within rounding error, each column totals 100%. The bolded row indicates the BARS dimension receiving a plurality of subject matter expert assignments.

Table 5 Initiative and Work Ethic

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
N_i	13	10	20	22	65
N_r	40	32	61	63	196
Communication skills (%)	3	0	0	0	1
Flexibility and resilience (%)	10	16	3	2	6
Initiative and work ethic (%)	**55**	**59**	**61**	**59**	**59**
Management/leadership (%)	5	9	2	0	3
Problem-solving skills (%)	8	0	2	3	3
Responsibility (%)	20	16	30	24	23
Service (%)	0	0	3	10	4
Teamwork and citizenship (%)	0	0	0	3	1

Note. N_i is the number of unique individuals who provided responses to the items chosen for the scale, whereas N_r reflects the number of behavioral statements rated by each individual. No item was answered more than once by the same individual. Percentages show the overall percentages of statements that were classified within each BARS dimension across all respondents (i.e., rates of respondent agreement that the pool of behavioral statements in question was aligned with each dimension). Within rounding error, each column totals 100%. The bolded row indicates the BARS dimension receiving a plurality of subject matter expert assignments.

Turning to the service and responsibility dimensions (Tables 4 and 9, respectively), our results indicate that responses from health care professionals differed somewhat from their peers' responses in other industries. Health care professionals showed the highest level of agreement with regard to items aligned with service (80%) compared to the other three industry subgroups (which ranged from 58% to 65%). While this may indicate that the service items were of particular salience to those in the health care industry, this subgroup was not distinguished as much from its peers with regard to ratings of behavioral statement relevancy (83% for health care vs. 74%, 82%, and 93% for administration, manufacturing, and human services, respectively). In the case of responsibility, 48% of responses from that subgroup agreed that the items dealt with the construct of responsibility, compared to the 73–84% agreement statistics among the other three industry subgroups. Although lower than the other industries, the 48% agreement figure from health care professionals regarding responsibility statements remained the highest level of dimensional agreement among those items. The second highest were assignments of these statements to initiative and work ethic, which, at 15% agreement, were far below the 48% figure for responsibility.

Table 6 Communication Skills

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
N_i	9	8	13	15	45
N_r	41	37	59	65	202
Communication skills (%)	**73**	**73**	**71**	**62**	**69**
Flexibility and resilience (%)	5	0	0	0	1
Initiative and work ethic (%)	0	0	3	0	1
Management/leadership (%)	7	3	3	9	6
Problem-solving skills (%)	5	8	3	2	4
Responsibility (%)	2	5	0	2	2
Service (%)	0	5	0	0	1
Teamwork and citizenship (%)	7	5	19	26	16

Note. N_i is the number of unique individuals who provided responses to the items chosen for the scale, whereas N_r reflects the number of behavioral statements rated by each individual. No item was answered more than once by the same individual. Percentages show the overall percentages of statements that were classified within each BARS dimension across all respondents (i.e., rates of respondent agreement that the pool of behavioral statements in question was aligned with each dimension). Within rounding error, each column totals 100%. The bolded row indicates the BARS dimension receiving a plurality of subject matter expert assignments.

Table 7 Flexibility and Resilience

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
N_i	13	10	20	22	65
N_r	38	26	61	65	190
Communication skills (%)	5	4	3	2	3
Flexibility and resilience (%)	**55**	**50**	**59**	**66**	**59**
Initiative and work ethic (%)	13	23	8	9	12
Management/leadership (%)	0	0	3	2	2
Problem-solving skills (%)	3	0	2	0	1
Responsibility (%)	21	8	13	17	15
Service (%)	0	0	0	0	0
Teamwork and citizenship (%)	3	15	11	5	8

Note. N_i is the number of unique individuals who provided responses to the items chosen for the scale, whereas N_r reflects the number of behavioral statements rated by each individual. No item was answered more than once by the same individual. Percentages show the overall percentages of statements that were classified within each BARS dimension across all respondents (i.e., rates of respondent agreement that the pool of behavioral statements in question was aligned with each dimension). Within rounding error, each column totals 100%. The bolded row indicates the BARS dimension receiving a plurality of subject matter expert assignments.

Mean effectiveness ratings for each subset of behavioral statements associated with the seven competency domains are displayed in Table 11. The tabled values reflect the effectiveness ratings for the behavioral statements that were used to represent the entire 1–6 effectiveness range for each of the BARS, and they exclude the statements that did not first meet the rule that the standard deviation of effectiveness ratings across raters should not exceed 1.39. This value was determined by adjusting the rule-of-thumb criterion of a standard deviation of 1.5 or below for 7-point scales (Pulakos, 2007) to take into account the 6-point rating scale SMEs used to evaluate statements' effectiveness in Study 2.[1] As part of the determination of generalizability of the BARS across industries, the intended focus of the tables should be on the consistency of means and standard deviations within rows of data. Overall and within industry subgroups, statements assigned to the problem-solving skills domain received higher effectiveness ratings on average than did statements in the other six competency areas. Interpreting between-area differences in such ratings is not advised, however, for two reasons. First, the group of statements being rated was mutually exclusive between domains (i.e., industry professionals were considering different material in providing ratings within each domain). Second, even if the rated statements *were*

Table 8 Problem-Solving Skills

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
N_i	8	5	13	15	41
N_r	16	11	25	31	83
Communication skills (%)	31	0	8	6	11
Flexibility and resilience (%)	0	9	0	3	2
Initiative and work ethic (%)	13	0	4	16	10
Management/leadership (%)	0	0	0	6	2
Problem-solving skills (%)	**44**	**73**	**80**	**55**	**63**
Responsibility (%)	6	0	4	3	4
Service (%)	0	0	0	0	0
Teamwork and citizenship (%)	6	18	4	10	8

Note. N_i is the number of unique individuals who provided responses to the items chosen for the scale, whereas N_r reflects the number of behavioral statements rated by each individual. No item was answered more than once by the same individual. Percentages show the overall percentages of statements that were classified within each BARS dimension across all respondents (i.e., rates of respondent agreement that the pool of behavioral statements in question was aligned with each dimension). Within rounding error, each column totals 100%. The bolded row indicates the BARS dimension receiving a plurality of subject matter expert assignments.

Table 9 Responsibility

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
N_i	13	10	20	22	65
N_r	47	33	75	79	234
Communication skills (%)	4	6	5	1	4
Flexibility and resilience (%)	4	12	1	3	4
Initiative and work ethic (%)	11	15	8	9	10
Management/leadership (%)	0	3	0	1	1
Problem-solving skills (%)	2	9	3	0	3
Responsibility (%)	**79**	**48**	**73**	**84**	**74**
Service (%)	0	0	4	3	2
Teamwork and citizenship (%)	0	6	5	0	3

Note. N_i is the number of unique individuals who provided responses to the items chosen for the scale, whereas N_r reflects the number of behavioral statements rated by each individual. No item was answered more than once by the same individual. Percentages show the overall percentages of statements that were classified within each BARS dimension across all respondents (i.e., rates of respondent agreement that the pool of behavioral statements in question was aligned with each dimension). Within rounding error, each column totals 100%. The bolded row indicates the BARS dimension receiving a plurality of subject matter expert assignments.

similar across domains, there would be no obvious standard by which to compare (for example) a rating of "effective" in one domain versus another. More helpful is to compare effectiveness ratings *within* a given domain across industry or gender subgroups. This is aided by the effect size statistics presented in Table 12. Considering industry subgroups, only nominal differences were exhibited across all subgroups in three of the seven competency areas (communication skills, flexibility and resilience, teamwork and citizenship). Service showed only one small difference in which $d = .20$, with initiative and work ethic showing two similarly sized small differences. The latter both involved ratings from health services professionals, who appeared to rate behavioral statements in this domain lower than their peers in both administration and health care. Responsibility exhibited three small mean differences, with the largest disparity ($d = .26$) observed between health care and health services workers. The greatest number and magnitude of mean differences in effectiveness ratings were observed with respect to problem-solving skills, in which two moderate effect sizes were evident (administration vs. health care and manufacturing, with $d = -.38$ and $d = -.37$, respectively).

Table 10 Teamwork and Citizenship

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
N_i	13	10	20	22	65
N_r	61	48	93	102	304
Communication skills (%)	5	4	9	4	6
Flexibility and resilience (%)	2	0	0	0	0
Initiative and work ethic (%)	7	2	3	2	3
Management/leadership (%)	7	8	8	11	9
Problem-solving skills (%)	0	8	1	2	2
Responsibility (%)	0	0	5	4	3
Service (%)	2	4	2	1	2
Teamwork and citizenship (%)	**79**	**73**	**72**	**76**	**75**

Note. N_i is the number of unique individuals who provided responses to the items chosen for the scale, whereas N_r reflects the number of behavioral statements rated by each individual. No item was answered more than once by the same individual. Percentages show the overall percentages of statements that were classified within each BARS dimension across all respondents (i.e., rates of respondent agreement that the pool of behavioral statements in question was aligned with each dimension). Within rounding error, each column totals 100%. The bolded row indicates the BARS dimension receiving a plurality of subject matter expert assignments.

Table 11 Effectiveness *N*s, Means, and Standard Deviations

	Industry				
	Administration	Health care	Human services	Manufacturing/ technology/construction	Total
Communication skills					
N	41	37	59	65	202
M (*SD*)	3.17 (1.99)	3.16 (1.98)	3.42 (2.06)	3.32 (1.95)	3.29 (1.98)
Flexibility and resilience					
N	38	26	61	65	190
M (*SD*)	3.71 (1.89)	3.65 (1.81)	3.57 (1.87)	3.78 (1.87)	3.68 (1.87)
Initiative and work ethic					
N	40	32	61	63	196
M (*SD*)	3.73 (1.87)	3.66 (1.91)	4.08 (1.72)	4.00 (1.65)	3.91 (1.75)
Problem-solving skills					
N	16	11	25	31	83
M (*SD*)	5.13 (1.02)	5.45 (0.69)	5.20 (1.41)	5.45 (0.72)	5.31 (1.02)
Responsibility					
N	47	33	75	79	234
M (*SD*)	3.36 (1.99)	3.73 (1.64)	3.24 (2.03)	3.33 (2.06)	3.36 (1.98)
Service					
N	39	30	60	66	195
M (*SD*)	4.08 (1.91)	3.97 (1.96)	3.73 (1.96)	3.68 (2.00)	3.82 (1.95)
Teamwork and citizenship					
N	61	48	93	102	304
M (*SD*)	3.51 (2.14)	3.63 (2.15)	3.60 (2.18)	3.44 (2.01)	3.53 (2.10)

Note. The tabled values reflect the effectiveness ratings for the behavioral statements that were used to represent the entire 1–6 effectiveness range, and they exclude the statements that did not first meet the standard deviation threshold of 1.39 (described in the main text). Therefore they are not subject to that standard deviation threshold.

We eliminated from further consideration all statements that exceeded the standard deviation threshold of 1.39 (48 out of 398 possible items, which represented about 12% of the statement pool). Statements that met the standard deviation criterion served as the pool of items used to construct the final BARS. Next, using effectiveness ratings and information about the relevance of the statements, R&D staff reviewed the statement pool and identified those statements that they determined would be optimal to form the final scales. Staff then met as a group to discuss these selections and indicate their

Table 12 Effectiveness Rating Effect Sizes Between Sector/Industry Subgroups

	A – HC	A – HS	A – M/T/C	HC – HS	HC – M/T/C	HS – M/T/C
Communication skills	.00	−.12	−.08	−.13	−.08	.05
Flexibility and resilience	.03	.07	−.04	.04	−.07	−.11
Initiative and work ethic	.04	−.20	−.16	−.23	−.19	.05
Problem-solving skills	−.38	−.06	−.37	.23	.00	−.22
Responsibility	−.20	.06	.02	.26	.21	−.04
Service	.06	.18	.20	.12	.14	.03
Teamwork and citizenship	−.05	−.04	.03	.01	.09	.08

Note. Figures are Cohen's *d* effect size statistics. A positive effect size indicates that the mean effect size rating for the first subgroup listed was greater than the mean for the second group listed. A = administration; HC = health care; HS = human services; M/T/C = manufacturing/technology/construction.

preliminary preferences. Two of the researchers made the final selection of statements. They ultimately selected statements based both on the recommendations of the group and the similarity of the behavioral content of the statements across the three effectiveness ranges. Statements were chosen such that there was rough continuity in the behavioral content running across the three effectiveness ranges. When necessary, statements were altered slightly to enhance this alignment. If the statement pool did not yield an adequate number of viable statements for an effectiveness range within a performance dimension, other statements of differing levels of effectiveness were modified to fill those gaps.

Three of the researchers reviewed these statements twice and, after discussion, came to consensus on their final forms. A final check on the instrument was provided by sending it to three external SMEs, all of whom have extensive publication records in highly regarded peer-reviewed industrial/organizational psychology journals. The job titles of the SMEs are Associate Professor of Personnel Management, Work, and Organizational Psychology; Professor of Organizational Psychology; and University Distinguished Professor of Organizational Psychology. All three SMEs reviewed the BARS favorably. The final scales appear in Appendix B, but the behavioral statements have been pixelated, as they are proprietary.

Discussion and Conclusion

To evaluate the performance of Zone Three jobs, we created seven BARS that capture 21st-century skills believed to be critical for success in modern Zone Three jobs. The BARS are a behaviorally based approach to employee assessment that provides the benefits of careful job analysis, including legal defensibility. With the help of supervisors of Zone Three workers, we developed behavioral statements for six BARS points. First, the managers generated approximately 430 critical incidents that we converted into behavioral statements. Next, a different set of managers substantiated that the statements were relevant for the Zone Three workers they oversaw. The managers retranslated statements back into dimensions, confirming that the statements reflected the appropriate dimensions. Moreover, the managers rated the effectiveness level for each behavioral statement. This process allowed us to anchor relevant statements to the appropriate scales at appropriate levels of effectiveness. Leading experts confirmed that the statements were appropriate for the scales that we were trying to develop, and then we finalized the BARS. But we did so only after the research literature and our own analyses confirmed that the BARS generalize across economic sectors. Organizations that seek to evaluate the performance of a wide variety of middle-skills employees in an economical way may find these BARS to be a useful tool.

This study has limitations, the major ones of which are mentioned here. The BARS dimensions, the ascription of the behavioral statements to dimensions, and statements' ratings (all of the foregoing being shown in Appendix A) might not generalize to all Zone Three jobs. Our sample is limited in size and scope. While this research provided evidence of the generalizability of findings across industries, we did not conduct any statistical significance testing or provide any confidence intervals to confirm that evidence. There are industries that we did not cover (e.g., agriculture). Moreover, industries can be broken up into subsectors, which might vary in terms of the applicability of the BARS. In addition, in terms of demographics (race, gender, etc.), our SMEs might not proportionally have represented the population of Zone Three managers. Also, our SME managers have observed behaviors of solely or mainly incumbent employees (rather than of all job applicants). The SMEs presumably did not observe work-related behaviors of those not hired. If the BARS are to be used for hiring, then the experiences on which the SMEs' responses were based to create the BARS are "range restricted" and thus might not be equivalently applicable across the entire applicant pool.

Furthermore, the standards used in the development of BARS (e.g., retranslation methods, the effectiveness ratings' standard deviation threshold for agreement, the number of scale points) are not inviolable. Different approaches might yield different conclusions. For example, a confirmatory factor analysis might yield a different conclusion than our retranslation efforts did in terms of the dimensions to which statements belonged. Additionally, some of the BARS items were derived from other BARS items rather than developed independently, a process that required additional researcher discretion.

Owing to the sensitivity of regression-based weighting to sample size and the number of independent variables, as well as the uncertainty of whether differential weighting would lead to more accurate decision-making than unit weighting would, we did not weight the dimensions to reflect greater importance that some of them may have for some job sectors or for some levels of job responsibility (e.g., mid-level vs. entry level). But it is possible that differential weighting might be beneficial. In summary, more research is needed. These limitations notwithstanding, this report provides evidence that a single set of BARS, with their psychometric, economic, and legal benefits, can be applied to evaluate middle-skills employees across multiple job sectors.

Acknowledgments

The authors would like to thank Richard J. Tannenbaum of Educational Testing Service for his technical and logistical guidance and support, which were instrumental to the success of this project. While they remain anonymous, the authors wish to acknowledge the generosity of assessment experts outside of Educational Testing Service for giving their time to review and comment upon behavioral statements. The authors are grateful as well to Adam N. Bacall of Educational Testing Service for his assistance with data analysis, editing, reformatting, and locating references. Furthermore, the authors thank Heather L. Walters (Educational Testing Service) for management of online surveys and Paola C. Heincke (Educational Testing Service) for helping them keep to certain timetables. In addition, the authors appreciate all of the helpful feedback from reviewers and editors that made this a better research report.

Note

1 $\sqrt{6/7 * 1.5^2} \approx 1.39$.

References

Allen, H. (2008). Using routinely collected data to augment the management of health and productivity loss. *Journal of Occupational and Environmental Medicine*, *50*, 615–632. https://doi.org/10.1097/jom.0b013e31817b610c

Allworth, E., & Hesketh, B. (1999). Construct-oriented biodata: Capturing change-related and contextually relevant future performance. *International Journal of Selection and Assessment*, *7*, 97–111. https://doi.org/10.1111/1468-2389.00110

Amini v. City of Minneapolis, 643 F.3d. 1068 (8th Cir. 2011).

Arvey, R. D., & Mussio, S. J. (1973). A test of expectancy theory in a field setting using female clerical employees. *Journal of Vocational Behavior*, *3*, 421–432. https://doi.org/10.1016/0001-8791(73)90054-7

Atkin, R. S., & Conlon, E. J. (1978). Behaviorally anchored rating scales: Some theoretical issues. *Academy of Management Review*, *3*, 119–128. https://doi.org/10.5465/amr.1978.4297058

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology*, *77*, 836–874. https://doi.org/10.1037//0021-9010.77.6.836

Bakan, D. (1966). The influence of phrenology on American psychology. *Journal of the History of the Behavioral Sciences*, *2*, 200–220. https://doi.org/10.1002/1520-6696(196607)2:3%3C200::aid-jhbs2300020304%3E3.0.co;2-l

Bakker, A. B., Demerouti, E., & Verbeke, W. (2004). Using the job demands-resources model to predict burnout and performance. *Human Resource Management*, *43*, 83–104. https://doi.org/10.1002/hrm.20004

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, *90*, 1185–1203. https://doi.org/10.1037/0021-9010.90.6.1185

Bartram, D., Robertson, I. T., & Callinan, M. (2002). Introduction: A framework for examining organizational effectiveness. In I. T. Robertson, M. Callinan, & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 1–10). Chichester, England: John Wiley. https://doi.org/10.1002/9780470696736.ch

Bernardin, H. J. (1977). Behavioral expectation scales versus summated scales: A fairer comparison. *Journal of Applied Psychology*, *62*, 422–427. https://doi.org/10.1037//0021-9010.62.4.422

Bernardin, H. J., LaShells, M. B., Smith, P. C., & Alvares, K. M. (1976). Behavioral expectation scales: Effects of developmental procedures and formats. *Journal of Applied Psychology*, *61*, 75–79. https://doi.org/10.1037//0021-9010.61.1.75

Bernardin, H. J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored ratings scales (BARS). *Journal of Applied Psychology*, *66*, 458–463. https://doi.org/10.1037//0021-9010.66.4.458

Blood, M. R. (1974). Spin-offs from behavioral expectation scale procedures. *Journal of Applied Psychology*, *59*, 513–515. https://doi.org/10.1037/h0037315

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, *64*, 410–421. https://doi.org/10.1037//0021-9010.64.4.410

Borman, W. C. (1986). Behavior-based rating scales. In R. A. Berk (Ed.), *Performance assessment: Methods and applications* (pp. 100–120). Baltimore, MD: The Johns Hopkins University Press.

Borman, W. C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. *2*, pp. 271–326). Palo Alto, CA: Consulting Psychologists Press. https://doi.org/10.1525/jung.1.1981.2.3.18

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, *6*, 1–21. https://doi.org/10.1207/s15327043hup0601_1

Borman, W. C., & Dunnette, M. D. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology*, *60*, 561–565. https://doi.org/10.1037//0021-9010.60.5.561

Borman, W. C., Hough, L. M., & Dunnette, M. D. (1976). *Development of behaviorally based rating scales for evaluating the performance of U.S. Navy recruiters* (Technical Report No. 76–31). San Diego, CA: Navy Personnel Research and Development Center. https://dx.doi.org/10.1037/e406422005-003

Borman, W. C., & Motowidlo, S. M. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco, CA: Jossey-Bass. https://doi.org/10.1207/s15327043hup1002_3

Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, *9*, 52–69. https://doi.org/10.1111/1468-2389.00163

Burrowes, J., Young, A., Restuccia, D., Fuller, J., & Raman, M. (2014, November). Bridging the gap: Rebuilding America's middle skills. https://dx.doi.org/10.1002/9781119171409.ch2

Burrus, J., Jackson, T., Xi, N., & Steinberg, J. (2013). *Identifying the most important 21st century workforce competencies: An analysis of the occupational Information network (O*NET)* (Research Report No. RR-13-21). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/9781119171409.ch2

Burton, W. N., Pransky, G., Conti, D. J., Chen, C. Y., & Edington, D. W. (2004). The association of medical conditions and presenteeism. *Journal of Occupational and Environmental Medicine*, *46*, S38–S45. https://doi.org/10.1097/01.jom.0000126687.49652.44

Campbell, C. H., Ford, P., Rumsey, M. G., Pulakos, E. D., Borman, W. C., Felker, D. B., … Riegelhaupt, B. J. (1990). Development of multiple job performance measures in a representative sample of jobs. *Personnel Psychology*, *43*, 277–300. https://doi.org/10.1111/j.1744-6570.1990.tb01559.x

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 687–732). Palo Alto, CA: Consulting Psychologists Press. https://doi.org/10.4135/9781473914940.n2

Campbell, J. P. (2012). Behavior, performance, and effectiveness in the twenty-first century. *Oxford handbook of organizational psychology* (Vol. *1*, pp. 159–194). Oxford, England: Oxford University Press. https://doi.org/10.1037/h0034185

Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. (1973). The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology*, *57*, 15–22. https://doi.org/10.1037/h0034185

Campbell, J. P., & Kuncel, N. R. (2011). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology: Vol. 1. Personnel psychology* (pp. 278–312). Thousand Oaks, CA: Sage. https://doi.org/10.1037/h0034185

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology*, *43*, 313–575. https://doi.org/10.1111/j.1744-6570.1990.tb01561.x

Carnevale, A. P., Gainer, L., & Meltzer, A. (1988). *Workforce basics: The skills employers want*. Washington, DC: U.S. Department of Labor & American Society for Training and Development. https://doi.org/10.1080/13678868.2013.821267

Cascio, W. F., & Aguinis, H. (2004). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Prentice Hall. https://doi.org/10.1108/ijm.2002.23.6.580.3

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century US workforce*. Washington, DC: Partnership for 21st Century Skills. https://doi.org/10.1007/978-94-6091-463-8_1

Chan, D. C. (2006). Core competencies and performance management in Canadian public libraries. *Library Management*, *27*, 144 – 153. https://doi.org/10.1108/01435120610652897

Coleman, V. I., & Borman, W. C. (2000). Investigating the underlying structure of the citizenship performance domain. *Human Resource Management Review*, *10*, 25 – 44. https://doi.org/10.1016/s1053-4822(99)00037-6

Engelbrecht, A. S., & Fischer, A. H. (1995). The managerial performance implications of a developmental assessment center process. *Human Relations*, *48*, 387 – 404. https://doi.org/10.1177/001872679504800405

Escorpizo, R. (2008). Understanding work productivity and its application to work-related musculoskeletal disorders. *International Journal of Industrial Ergonomics*, *38*, 291 – 297. https://doi.org/10.1016/j.ergon.2007.10.018

Farr, J. L., & Levy, P. E. (2007). Performance appraisal. In L. L. Koppes (Ed.), *Historical perspectives in industrial and organizational psychology* (pp. 311 – 327). Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.1002/jhbs.20264, *43*

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327 – 358.https://doi.org/10.1037/h0061470

Fluegge, E. R. (2008). Who put the fun in functional? Fun at work and its effects on job performance. *Dissertation Abstracts International*, *69*, 2781A. https://doi.org/10.1108/mrr-11-2012-0252

Freyd, M. (1923). The graphic rating scale. *Journal of Educational Psychology*, *14*, 83 – 102. https://doi.org/10.1037/h0074329

Golubovich, J., Su, R., & Robbins, S. (2017). Establishing an international standards framework and action research agenda for workplace readiness and success. In J. Burrus, K. Mattern, B. Naemi, & R. Roberts (Eds.), *Building better students: Preparation for the workforce* (pp. 303 – 337). New York, NY: Oxford University Press. https://doi.org/10.1037/h0074329, *14*

Goodale, J. G., & Burke, R. J. (1975). Behaviorally based rating scales need not be job specific. *Journal of Applied Psychology*, *60*, 389 – 391. https://doi.org/10.1037/h0076629

Greenslade, J. H., & Jimmieson, N. L. (2007). Distinguishing between task and contextual performance for nurses: Development of a job performance scale. *Journal of Advanced Nursing*, *58*, 602 – 611. https://doi.org/10.1111/j.1365-2648.2007.04256.x

Griffin, B., & Hesketh, B. (2003). Adaptable behaviours for successful work and career adjustment. *Australian Journal of Psychology*, *55*, 65 – 73. https://doi.org/10.1080/00049530412331312914

Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal*, *50*, 327 – 347.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill. https://doi.org/10.5465/AMJ.2007.24634438

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). New York, NY: Routledge. https://doi.org/10.5465/AMJ.2006.20786077

Harrison, D. A., Newman, D. A., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy of Management Journal*, *49*, 305 – 325. https://doi.org/10.5465/AMJ.2006.20786077

Hedge, J. W., Borman, W. C., Bruskiewicz, K. T., & Bourne, M. J. (2004). The development of an integrated performance category system for supervisory jobs in the US Navy. *Military Psychology*, *16*, 231 – 243. https://doi.org/10.1207/s15327876mp1604_2

Hom, P. W., DeNisi, A. S., Kinicki, A. J., & Bannister, B. D. (1982). Effectiveness of performance feedback from behaviorally anchored rating scales. *Journal of Applied Psychology*, *67*, 568 – 576. https://doi.org/10.1037/0021-9010.67.5.568

Hunt, S. T. (1996). Generic work behavior: An investigation into the dimensions of entry-level, hourly job performance. *Personnel Psychology*, *49*, 51 – 83. https://doi.org/10.1111/j.1744-6570.1996.tb01791.x

Ivancevich, J. M. (1980). A longitudinal study of behavioral expectation scales: Attitudes and performance. *Journal of Applied Psychology*, *65*, 139 – 146. https://doi.org/10.1037/0021-9010.65.2.139

Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. *Personnel Psychology*, *33*, 595 – 640. https://doi.org/10.1111/j.1744-6570.1980.tb00486.x

Jeanneret, P. R., & Zedeck, S. (2010). Professional guidelines/standards. In J. Farr & N. Tippins, (Eds.), *Handbook of employee selection* (pp. 593 – 625). New York, NY: Taylor & Francis.

Jiambalvo, J. (1979). Performance evaluation and directed job effort: Model development and analysis in a CPA firm setting. *Journal of Accounting Research*, *17*, 436 – 455. https://doi.org/10.2307/2490512

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits — Self-esteem, generalized self-efficacy, locus of control, and emotional stability — With job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*, 80 – 92 https://doi.org/10.1037/0021-9010.86.1.80

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, *87*, 797 – 807. https://doi.org/10.1037/0021-9010.87.4.797

Keaveny, T. J., & McGann, A. F. (1975). A comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology*, *60*, 695 – 703. https://doi.org/10.1037/0021-9010.60.6.695

Kell, H. J., Martin-Raugh, M. P., Carney, L. M., Inglese, P. A., Chen, L., & Feng, G. (2017). *Exploring methods for developing behaviorally-anchored scales for evaluating structured interview performance* (Research Report No. RR-17-28). Princeton, NJ: Educational Testing Service. https://.org/10.1002/ets2.12152

Kemeny, T., & Rigby, D. (2012). Trading away what kind of jobs? Globalization, trade and tasks in the US economy. *Review of World Economics*, *148*, 1–16. https://doi.org/10.1007/s10290-011-0099-5

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Schaufeli, W. B., de Vet, H. C. W., & van der Beek, A. J. (2011). Conceptual frameworks of individual work performance: A systematic review. *Journal of Occupational and Environmental Medicine*, *53*, 856–866. https://doi.org/10.1097/JOM.0b013e318226a763

Kurz, R., & Bartram, D. (2002). Competency and individual performance: Modeling the world of work. In I. T. Robertson, M. Callinan, & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 227–255). Chichester, England: John Wiley. https://doi.org/10.1002/9780470696736.ch10

Kyllonen, P. C. (2016). Socio-emotional and self-management variables in learning and assessment. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 174–197). New York, NY: John Wiley. https://doi.org/10.1002/9781118956588.ch8

Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. *Journal of Applied Psychology*, *77*, 437–452. https://doi.org/10.1037/0021-9010.77.4.437

Landy, F., Gutman, A., & Outtz, J. (2010). A sampler of legal principles in employment selection. In J. Farr & N. Tippins (Eds.), *Handbook of employee selection* (pp. 627–676). New York, NY: Taylor & Francis.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, *87*, 72–107. https://doi.org/10.1037/0033-2909.87.1.72

Landy, F. J., & Guion, R. M. (1970). Development of scales for the measurement of work motivation. *Organizational Behavior and Human Performance*, *5*, 93–103. https://doi.org/10.1016/0030-5073(70)90007-3

Latham, G. P., Fay, C. H., & Saari, L. M. (1979). The development of behavioral observation scales for appraising the performance of foremen. *Personnel Psychology*, *32*, 299–311. https://doi.org/10.1111/j.1744-6570.1979.tb02136.x

Lee, C., & Tindal, G. (1996). *Construction and use of behavior rating scales*. Eugene, OR: University of Oregon, Research, Consultation, & Teaching Program.

LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, *87*, 52–65. https://doi.org/10.1037/0021-9010.87.1.52

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*, 241–293. https://doi.org/10.1111/peps.12052

Luo, Z. X., Shi, K., Li, W. D., & Miao, D. M. (2008). Construct of job performance: Evidence from Chinese military soldiers. *Asian Journal of Social Psychology*, *11*, 222–231. https://doi.org/10.1111/j.1467-839X.2008.00261.x

Mael, F. A., O'Shea, P. G., Smith, M. A., Burling, A. S., Carman, K. L., Haas, A., & Rogers, K. S. (2010). Development of a model and measure of process-oriented quality of care for substance abuse treatment. *Journal of Behavioral Health Services and Research*, *37*, 4–24. https://doi.org/10.1007/s11414-009-9180-4

Martin-Raugh, M., Tannenbaum, R. J., Tocci, C. M., & Reese, C. (2016). Behaviorally anchored rating scales: An application for evaluating teaching practice. *Teaching and Teacher Education*, *59*, 414–419. https://doi.org/10.1016/j.tate.2016.07.026

Maxham, J. G., III, Netemeyer, R. G., & Lichtenstein, D. R. (2008). The retail value chain: Linking employee perceptions to employee performance, customer evaluations, and store performance. *Marketing Science*, *27*, 147–167. https://doi.org/10.1287/mksc.1070.0282

McReynolds, P., & Ludwig, K. (1987). On the history of rating scales. *Personality and Individual Differences*, *8*, 281–283. https://doi.org/10.1016/0191-8869(87)90188-7

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: American Council of Education and Macmillan.

Michel, R. P. (2000). A model of entry-level job performance. *Dissertation Abstracts International*, *61*, 3313B.

Miner, J. B. (1917). The evaluation of a method for finely graduated estimates of abilities. *Journal of Applied Psychology*, *1*, 123–133. https://doi.org/10.1037/h0071226

Motowidlo, S. J., & Kell, H. J. (2013). Job performance. In N. W. Schmitt & S. Highhouse (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 82–103). Hoboken, NJ: John Wiley.

Muck, P. M., Hell, B., & Höft, S. (2008). Application of the principles of behaviorally anchored rating scales to assess the big five personality constructs at work. In J. Deller (Ed.), *Research contributions to personality at work* (pp. 77–97). Munich, Germany: Rainer Hampp.

Murphy, K. R. (1989). Dimensions of job performance. In R. F. Dillon & J. W. Pellegrino (Eds.), *Testing: Theoretical and applied perspectives* (pp. 218–247). New York, NY: Praeger.

Nathan, B. R., & Cascio, W. F. (1986). Technical and legal standards. In R. A. Berk (Ed.), *Performance assessment: Methods and application* (pp. 1–50). Baltimore, MD: The Johns Hopkins University Press.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13398

O*NET Online. (n.d.). *O*NET online help: Job zones*. Retrieved from http://www.onetonline.org/help/online/zones

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., … Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning & Education*, *11*, 609 – 630. https://doi.org/10.5465/amle.2010.0177

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books/DC Heath and Com.

Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, *10*, 85 – 97. https://doi.org/10.1207/s15327043hup1002_2

Paterson, D. G. (1922). The Scott company graphic rating scale. *Journal of Personnel Research*, *1*, 361 – 376.

Paterson, D. G. (1923). Methods of rating human qualities. *Annals of the American Academy of Political and Social Science*, *110*, 81 – 93. https://doi.org/10.1177/000271622311000109

Pierce v. Board of Regents of the University System of Georgia (M.D. Ga. 2011).

Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly*, *1*, 107 – 142. https://doi.org/10.1016/1048-9843(90)90009-7

Pulakos, E. D. (2007). Performance measurement. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement* (pp. 293 – 318). Mahwah, NJ: Lawrence Erlbaum.

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, *85*, 612 – 624.https://doi.org/10.1037/0021-9010.85.4.612

Pulakos, E. D., Schmitt, N., Dorsey, D. W., Arad, S., Hedge, J. W. & Borman, W. C., (2002). Predicting adaptive performance: Further tests of a model of adaptability. *Human Performance*, *15*, 299 – 323. https://doi.org/10.1207/S15327043HUP1504_01

Reilly, N. P., Bocketti, S. P., Maser, S. A., & Wennet, C. L. (2006). Benchmarks affect perceptions of prior disability in a structured interview. *Journal of Business and Psychology*, *20*, 489 – 500. https://doi.org/10.1007/s10869-005-9005-2

Renn, R. W., & Fedor, D. B. (2001). Development and field test of a feedback seeking, self-efficacy, and goal setting model of work performance. *Journal of Management*, *27*, 563 – 583. https://doi.org/10.1177/014920630102700504

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skills factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*, 261 – 288. https://doi.org/10.1037/0033-2909.130.2.261

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*, *38*, 555 – 572.

Robinson, S. L., & O'Leary-Kelly, A. M. (1998). Monkey see, monkey do: The influence of work groups on the antisocial behavior of employees. *Academy of Management Journal*, *41*, 658 – 672. https://doi.org/10.2307/256693

Roch, S. G., Sternburgh, A. M., & Caputo, P. M. (2007). Absolute vs relative performance rating formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment*, *15*, 302 – 316. https://doi.org/10.1111/j.1468-2389.2007.00390.x

Rollins, T., & Fruge, M. (1992). Performance dimensions: Competencies with a twist. *Training*, *29*(1), 47 – 51.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*, 66 – 80. https://doi.org/10.1037/0021-9010.87.1.66

Rumsey, M. G., & Mietus, J. R. (1983). Psychometric comparison of two alternative rating scale formats. In J. E. Woolam (Ed.), *Proceedings of the 25th annual conference of the military testing association* (pp. 461 – 466). Pensacola, FL: Naval Education and Training Program Development Center.

Sackett, P. R. (2002). The structure of counterproductive work behaviors: Dimensionality and relationships with facets of job performance. *International Journal of Selection and Assessment*, *10*, 5 – 11. https://doi.org/10.1111/1468-2389.00189

Schwab, D. P., Heneman, H. G., & DeCotiis, T. A. (1975). Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology*, *28*, 549 – 562. https://doi.org/10.1111/j.1744-6570.1975.tb01392.x

Shore, J., Lentini, J., Rikoon, S., Seybert, J., & Noeth, R. (2016). The ETS work readiness strength assessment and training system *(ETS product description)*. Princeton, NJ: Educational Testing Service.

Sinclair, R. R., & Tucker, J. S. (2006). Stress-CARE: An integrated model of individual differences in soldier performance under stress. In T. W. Britt, C. A. Castro, & A. B. Adler (Eds.), *Military life: The psychology of serving in peace and combat: Vol. 1. Military performance* (pp. 202 – 231). Westport, CT: Praeger Security International.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, *47*, 149 – 155. https://doi.org/10.1037/h0047060

Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and content validation of a "hyperdimensional" taxonomy of managerial competence. *Human Performance*, *13*, 205 – 251. https://doi.org/10.1207/S15327043HUP1303_1

Tziner, A. (1984). A fairer examination of rating scales when used for performance appraisal in a real organizational setting. *Journal of Organizational Behavior*, *5*, 103 – 112. https://doi.org/10.1002/job.4030050203

Tziner, A., & Kopelman, R. E. (2002). Is there a preferred performance rating format? A non-psychometric perspective. *Applied Psychology*, *51*, 479 – 503. https://doi.org/10.1111/1464-0597.00104

U.S. Department of Labor. (2014). *WIOA overview*. Retrieved from https://www.doleta.gov/wioa/Overview.cfm

U.S. Department of Labor. (2015). *Labor force statistics from the current population survey. Household data annual averages 7. Employment status of the civilian noninstitutional population 25 years and over by educational attainment, sex, race, and Hispanic or Latino ethnicity*. Retrieved from http://www.bls.gov/cps/cpsaat07.htm

Van Dyne, L., Jehn, K. A., & Cummings, A. (2002). Differential effects of strain on two forms of work performance: Individual employee sales and creativity. *Journal of Organizational Behavior*, *23*, 57 – 74. https://doi.org/10.1002/job.127

Viswesvaran, C. (1993). *Modeling job performance: Is there a general factor?* (Unpublished doctoral dissertation). University of Iowa, Iowa City.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8*, 216 – 226. https://doi.org/10.1111/1468-2389.00151

Warner, L., Gates, S. L., Christeson, W., & Kiernan, M. (2011). *The skills gap: Reversing Washington's lack of skilled workers through early learning*. Retrieved from http://cdn.americasedge.org/clips/WA-Skills-Report.pdf

Wisecarver, M. M., Carpenter, T. D., & Kilcullen, R. N. (2007). Capturing interpersonal performance in a latent performance model. *Military Psychology*, *19*, 83 – 101. https://doi.org/10.1080/08995600701323376

Appendix A

Job Performance Taxonomies

Table A1

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
	Definition	Conducting oneself in a polite, patient, cooperative manner with individuals external to the organization (e.g., clients, customers, inspectors/auditors, vendors); acting to meet these individuals' needs; following through with these individuals to get the job done well; appropriately managing these individuals if they become difficult	Anticipating problems and solving them before they arise; persisting in difficult, long-term, or unpleasant tasks until they are successfully accomplished; completing tasks effectively (e.g., accurately, efficiently, in a timely manner)	Speaking, writing, and listening effectively; understanding and appropriately responding to people's actions in face-to-face communications, by email, teleconferences, and/or video conferences; reading others' body language	Adapting well to unclear or changing work demands; handling stress appropriately; accepting criticism or feedback without getting discouraged or overly defensive; remaining focused on getting the job done when faced with challenges and setbacks	Using information, knowledge, and reasoning to solve problems; thinking critically and creatively; using good judgment when making decisions; looking for new information, when necessary, to solve problems	Being accountable for one's own duties and actions; following safety and other rules, procedures, and policies; maintaining high standards of personal conduct and professionalism (e.g., being ethical, respectful)	Working well with all members of the organization, both individually and in groups; demonstrating respect for different opinions, customs, and preferences; actively participating in formal and informal group processes; being cooperative, helpful, and supportive to others	Leading, directing, mentoring, evaluating, giving feedback, and setting work-related goals and expectations for others (note: does not have to be an official requirement of one's job)	
Citation	**Methodology**									
Allen (2008)	17,821 health risk appraisal surveys were analyzed using structural equation modeling						Absenteeism Presenteeism			

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communi-cation skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Allworth and Hesketh (1999)	Task, contextual and adaptive measures of performance were used as criterion data in a selection study involving 325 staff in the hotel industry. Construct-oriented biodata were developed to predict adaptive and contextual performance and the validities were contrasted with measures of cognitive ability and personality		Task performance; contextual performance		Adaptive performance					
Arvey and Mussio (1973)	266 female clerical workers were rated on seven dimensions of job performance and surveyed. *not clear what kind of analysis*	Dealing with public	Showing concern for time; working accurately; detail and planning				Cooperating and extra time; showing responsibility and initiative	Cooperating and extra time; dealing with others in organization	Showing responsibility and initiative	Showing responsibility and initiative

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Bakker et al. (2004)	146 surveys from individuals from diverse fields including industrial work, business, transportation, education, and health care were analyzed using structural equation modeling		In-role performance; extra-role performance							
Bartram et al. (2002); Kurz and Bartram (2002); Bartram (2005)	Loadings of 112 component competencies on eight factors	Organizing and executing	Enterprising and performing	Interacting and presenting	Creating and conceptualizing; adapting and coping	Analyzing and interpreting	Organizing and executing	Supporting and cooperating	Leading and deciding	Leading and deciding
Borman and Brush (1993)	Critical incidents analysis, factor analysis		Technical activities	Interpersonal dealings and communication			Useful personal behavior and skills	Interpersonal dealings and communication; useful personal behavior and skills	Mechanics of management; leadership and supervision	Mechanics of management; leadership and supervision
Borman and Motowidlo (1993)	Review of existing literature		Task performance; contextual performance							
Burrus et al. (2013)	Descriptive stats on the importance scale ratings for the O*NET knowledge, skills, abilities, and work style domains for Job Zones 3		Dependability; attention to detail; initiative; independence; persistence		Stress tolerance; adaptability/flexibility		Integrity; self-control	Cooperation		

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Burton et al. (2004)	16,651 employees were surveyed from large financial services corporation. Multivariate logistic regression was used to uncover relationships between variables.						Absenteeism Presenteeism			
Campbell, Ford, et al. (1990)	Task analyses, critical incident analysis, factor analysis		Job-specific proficiency; non-job-specific proficiency							
Campbell (1990)	Review of existing literature, including interview and critical incident methods. Content analysis and job analysis was also employed.		Job-specific task proficiency; demonstrating effort	Written and oral communications			Maintaining personal discipline	Facilitating peer and team performance	Supervision and leadership; management and administration	Supervision and leadership; management and administration
Campbell (2012)	Review of existing literature		Technical performance; initiative, persistence and effort	Communication			Absence of counterproductive work behavior	Peer/team member leadership performance	Supervisory, manager, executive (i.e., hierarchical) leadership; management performance; peer/team member management performance	Supervisory, manager, executive (i.e., hierarchical) leadership; management performance; peer/team member management performance

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Campbell, McHenry, and Wise (1990)	Content analysis, confirmatory factor analysis		Core technical proficiency; general soldiering proficiency; physical fitness and military bearing; Effort				Personal discipline		Leadership	Leadership
Casner-Lotto and Barrington (2006)	Surveyed employers about skills needed for new hires		Professionalism; work ethic	Oral and written communication skills		Critical thinking; problem solving; innovation; creativity	Ethics; social responsibility	Teamwork and collaboration	Leadership	
Chan (2006)	Interviews	Customer service		Communication skills		Analytical skills		Interpersonal skills		
Engelbrecht and Fischer (1995)	BARS and factor analysis		Action orientation; development			Probing, synthesis and judgment	Empathy	Empathy	Managing information; task structuring	Managing information; task structuring
Escorpizo (2008)	Review of existing literature						Absenteeism Presenteeism			
Fluegge (2008)	Principal components analysis, confirmatory factor analysis, usefulness analysis, and structural equation modeling techniques		Task performance; organizational citizenship behavior			Creative performance				
Golubovich et al. (2017)	Review of existing literature, factor analysis.	Customer service	General task performance; proactive work behavior				Absence of counterproductive work behavior; safety and rule compliance	Teamwork		
Greenslade and Jimmieson (2007)	Exploratory factor analysis, validity testing		Task performance; contextual performance							

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Griffin et al. (2007)	Supervisor ratings and employee self-ratings were used to explore the factor structure of items assessing nine dimensions of workplace performance		Task proficiency; proactivity		Adaptability					
Hedge et al. (2004)	Critical incidents analysis, principal components analysis		Organizational savvy; personal and professional development	Communication skills			Professionalism and integrity		Coaching and mentoring; leading change; leading people; resource stewardship	Coaching and mentoring; leading change; leading people; resource stewardship
Hunt (1996)	Review of existing literature; analysis of supervisor ratings of specific employee behaviors gathered from 18,146 employees in 42 different hourly, entry-level jobs in predominantly retail settings		Industriousness; thoroughness; attendance		Schedule flexibility		Lack of off-task behavior, unruliness, theft, drug misuse; adherence to rules			
Jiambalvo (1979)	Multiple regression and correlation analyses					Understanding; planning; revising; recognizing problems; suggesting solutions		Cooperation; respect; promoting	Providing training; reviewing work; recognizing problems; promoting	Providing training; reviewing work; recognizing problems; promoting

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Lance et al. (1992)	Hierarchical confirmatory factor analysis		Technical proficiency					Interpersonal proficiency		
Luo et al. (2008)	Critical incidents analysis, confirmatory factor analysis, regression analysis		Task accomplishment; work capability; promoting organizational benefit; military training; love of learning				Self-discipline	Helping others	Helping others	Helping others
Mael et al. (2010)	*No information available*		Providing clinical services; clinical support; employee citizenship behavior						Managerial behavior	Managerial behavior
Maxham III et al. (2008)	Hierarchical linear modeling and structural equation path modeling	Extra-role performance toward customers	In-role performance; extra-role performance toward organization							
Michel (2000)	Unavailable		Task performance; civic performance					Interpersonal performance		
Murphy (1989)	Review of existing literature.		Task behaviors				Absence of downtime behaviors, destructive/hazardous behaviors	Interpersonal behaviors		

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communi-cation skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Pulakos, Schmitt, Dorsey, Arad, Hedge, & Borman, (2002)	739 military personnel completed three adaptability measures as well as cognitive and noncognitive assessments. Supervisors rated participants. Confirmatory factor analysis and item-parceling strategywas conducted		Task performance; contextual performance		Adaptive performance					
Renn and Fedor (2001)	Confirmatory factor analysis with survey data and employee records/ratings for 150 employees from customer sales/service representatives		Work quantity and quality							
Rollins and Fruge (1992)	*No information available*		Task proficiency; development	Communication	Creativity; adaptability		Accountability	Teamwork	Leadership; decision-making	Leadership; decision-making

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Rotundo and Sackett (2002)	Within-subjects regression analyses, hierarchical linear modeling, and hierarchical cluster analyses with surveys from 504 employees from accountancy, nursing, administrative, retail, and machine operator fields		Task performance; organizational citizenship behavior				Absence of counterproductive behavior			
Shore et al. (2016)	Review of existing literature	Friendliness, generosity, collaboration	Dependability, diligence, assertiveness		Stability, optimism	Intellectual orientation, creativity, inquisitiveness	Organization, self-discipline, dependability	Generosity, collaboration		
Sinclair and Tucker (2006)	*Insufficient information available*		Task performance; contextual performance		Adaptive performance		Lack of counterproductive behavior			
Tett et al. (2000)	Review of existing literature, classification of behaviors by SMEs	Person orientation	Traditional functions; task orientation; dependability; developing of self and others; occupational acumen and concerns	Communication	Open mindedness		Emotional control	Person orientation; developing of self and others	Person orientation; developing of self and others	Person orientation; developing of self and others
Van Dyne et al. (2002)	Principal component factor analysis		Sales performance			Creativity				

Table A1 Continued

	Dimension	Service (if applicable)	Initiative and work ethic	Communication skills	Flexibility and resilience	Problem-solving skills	Responsibility	Teamwork and citizenship	Leadership	Management (for managers only)
Viswesvaran (1993)	Lexical hypothesis		Overall work performance; productivity; quality; job knowledge; effort	Communication competence			Compliance with and acceptance of authority	Interpersonal competence	Leadership; administrative competence	Leadership; administrative competence
Viswesvaran and Ones (2000)	Review of existing literature	Task performance					Absence of counterproductive behavior			
Warner et al. (2011)			Attention to detail; initiative, perseverance, and independence	Communication	Personal flexibility	Mechanical reasoning, logic, trouble-shooting, & spatial visualization	Self-control and dependability; making independent decisions	Cooperation		
Wisecarver et al. (2007)	Confirmatory factor analysis		Job-specific task proficiency; interpersonal job-specific task proficiency; non-job-specific task proficiency; effort				Discipline	Interpersonal job-specific task proficiency; peer-team interaction	Management	Management

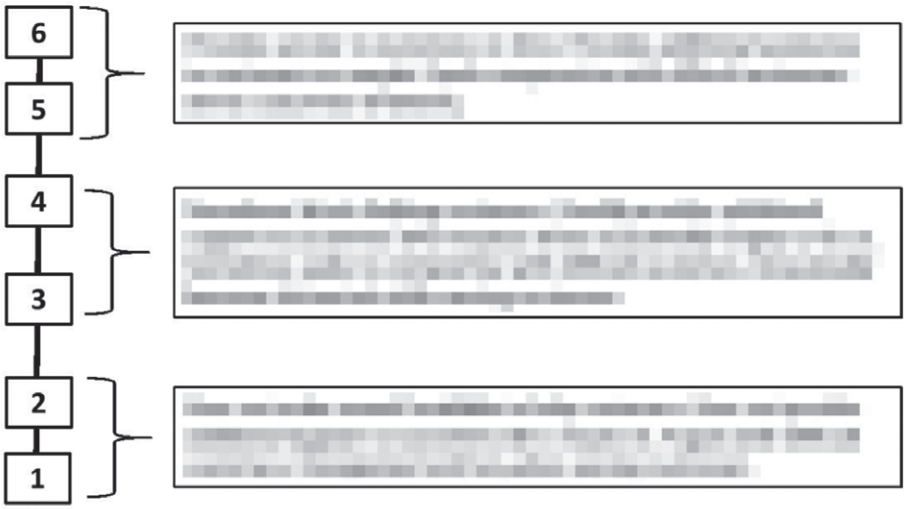
Appendix B

BARS Scales

The statements in the scales have been pixelated, because they are proprietary.

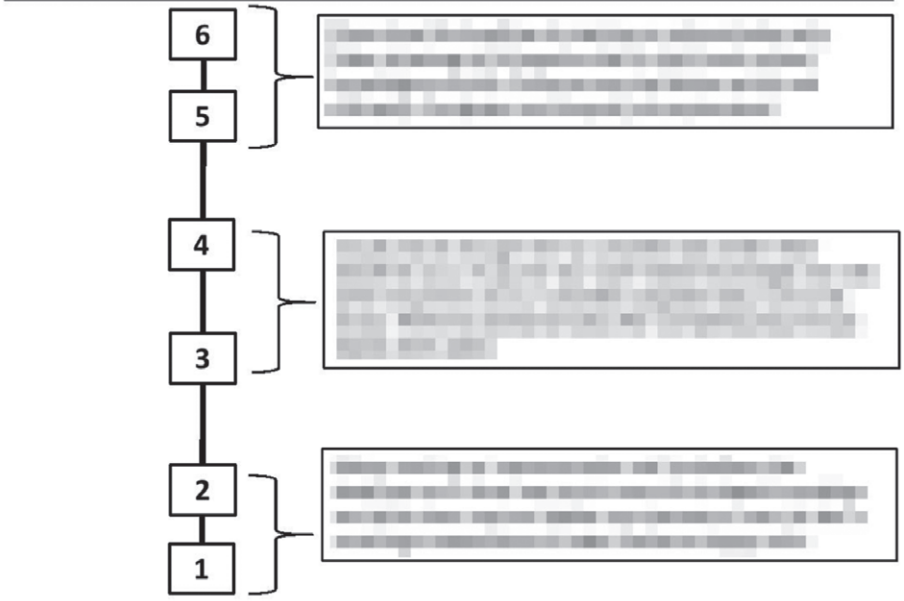
Service

Conducting oneself in a polite, patient, cooperative manner with individuals external to the organization (e.g. clients, customers, inspectors/auditors, vendors); acting to meet these individuals' needs; following through with these individuals to get the job done well; appropriately managing these individuals if they become difficult.



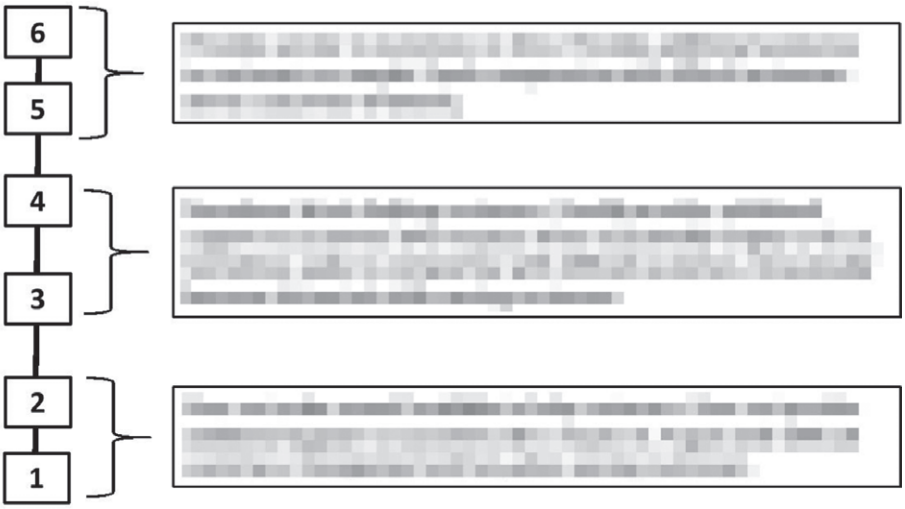
Initiative and Work Ethic

Anticipating problems and solving them before they arise; persisting in difficult, long-term, or unpleasant tasks until they are successfully accomplished; completing tasks effectively (e.g. accurately, efficiently, in a timely manner).



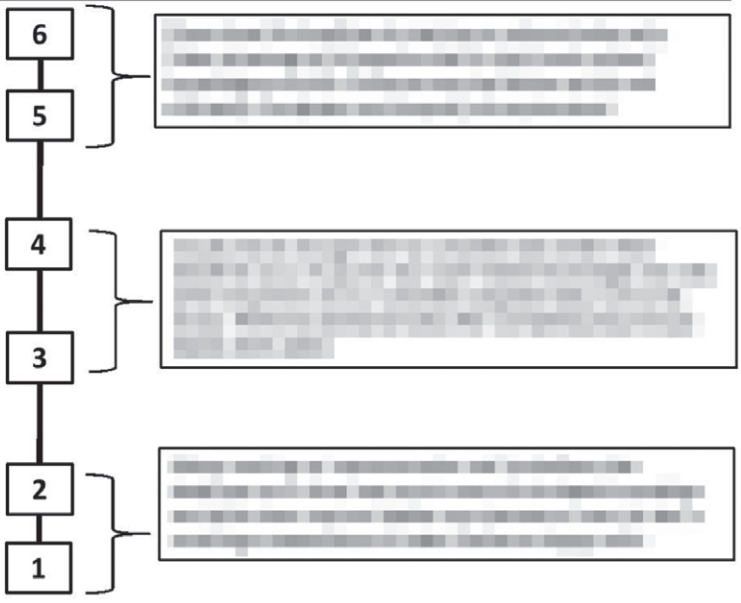
Service

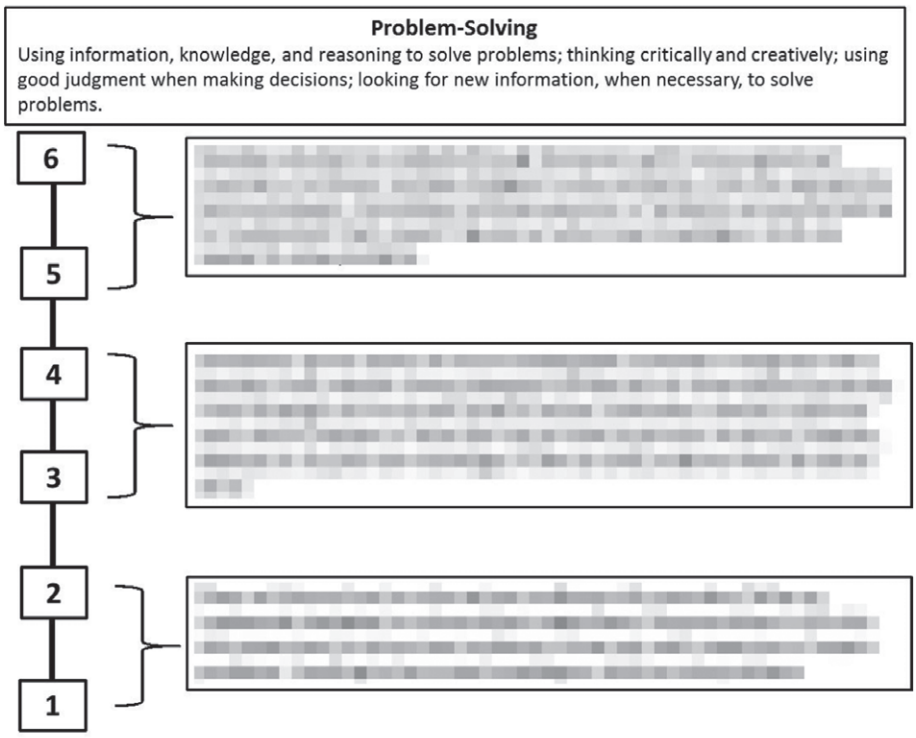
Conducting oneself in a polite, patient, cooperative manner with individuals external to the organization (e.g. clients, customers, inspectors/auditors, vendors); acting to meet these individuals' needs; following through with these individuals to get the job done well; appropriately managing these individuals if they become difficult.

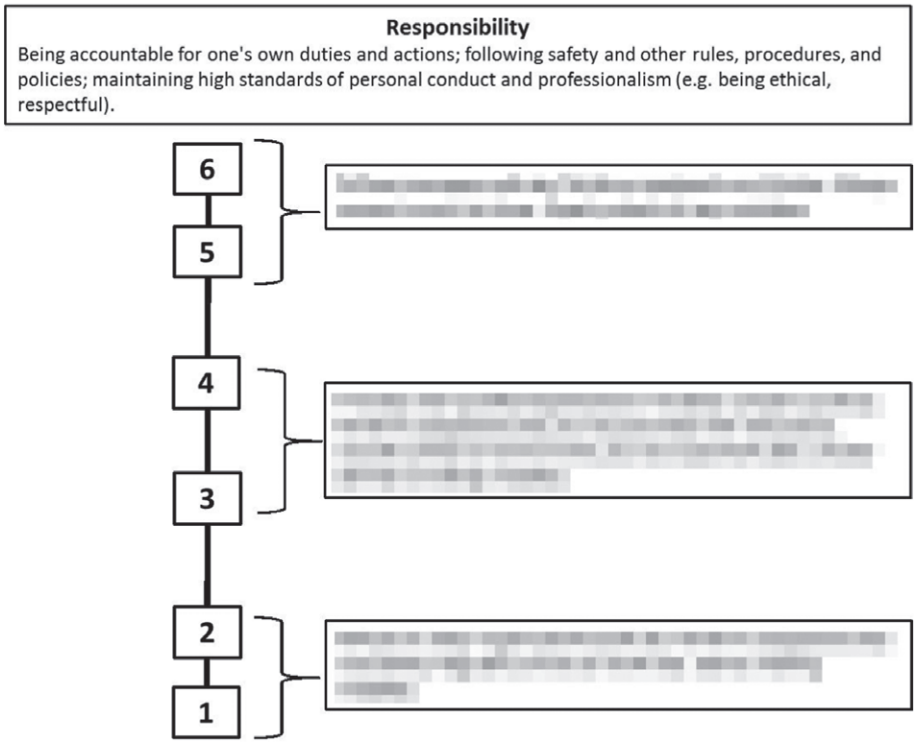


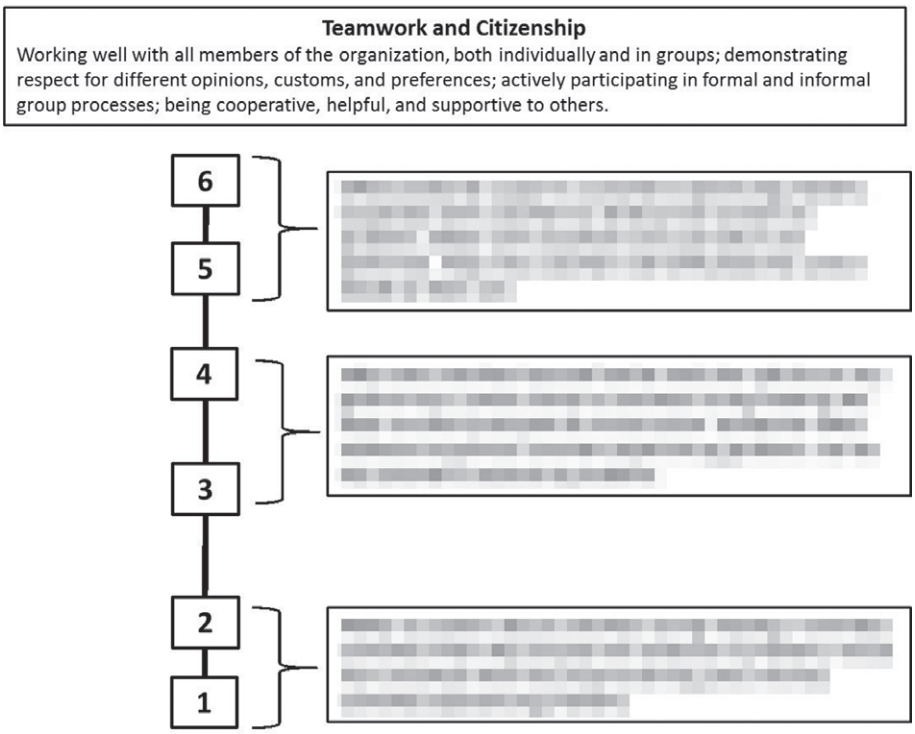
Initiative and Work Ethic

Anticipating problems and solving them before they arise; persisting in difficult, long-term, or unpleasant tasks until they are successfully accomplished; completing tasks effectively (e.g. accurately, efficiently, in a timely manner).




Problem-Solving
Using information, knowledge, and reasoning to solve problems; thinking critically and creatively; using good judgment when making decisions; looking for new information, when necessary, to solve problems.
6
5
4
3
2
1


Responsibility
Being accountable for one's own duties and actions; following safety and other rules, procedures, and policies; maintaining high standards of personal conduct and professionalism (e.g. being ethical, respectful).
6
5
4
3
2
1



Suggested citation:

Klieger, D. M., Kell, H. J., Rikoon, S., Burkander, K. N., Bochenek, J. L., & Shore, J. R. (2018). *Development of the behaviorally anchored rating scales for the skills demonstration and progression guide* (Research Report No. RR-18-24). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12210

Action Editor: James Carlson

Reviewers: Priya Kannan and Patrick Kyllonen