EFFECTS OF DESCRIPTOR SPECIFICITY AND OBSERVABILITY ON INCUMBENT WORK ANALYSIS RATINGS

ERICH C. DIERDORFF
Kellstadt Graduate School of Business
Department of Management
DePaul University

FREDERICK P. MORGESON
Eli Broad Graduate School of Management
Department of Management
Michigan State University

Judgments regarding the requirements of jobs and the requirements of individuals performing those jobs comprise the critical groundwork on which human resource practices are built. Yet, such judgments are potentially limited in a variety of ways. Using a large sample of incumbents ($N = 47,137$) spanning over 300 different occupations, we present research empirically examining how the specificity and observability of focal descriptors affect subsequent ratings. We use variance component (VC) estimation and meta-analysis to investigate sources of variance and interrater reliability of importance ratings across 5 descriptors (tasks, responsibilities, knowledge, skills, and traits). Results indicate that when ratings are rendered on descriptors of low specificity and low observability (e.g., traits), variance due to rater idiosyncrasies increases and reliability decreases. Implications for work analysis practice and future research are discussed.

Job or work analysis forms the foundation for virtually every human resource (HR) management system (Butler & Harvey, 1988). For example, work analysis serves as the foundation for recruiting and selecting workers, designing and redesigning work, developing training programs, determining the level of compensation to provide, and designing performance management systems. To develop these HR systems, information about a variety of different job and worker requirements must be collected. These requirements would include descriptive information about the tasks performed on the job, the general responsibilities of incumbents, and the different types of knowledge, skill, ability, and other characteristics that are needed to perform the job (Dierdorff & Morgeson, 2007). The determination of these requirements, however, is largely based on human

---

judgment (Goldstein, Zedeck, & Schneider, 1993), which is thought to be limited in numerous ways (Morgeson & Campion, 1997).

Although research has begun to investigate differences in job and worker requirements (Dierdorff & Morgeson, 2007; Dierdorff & Wilson, 2003; Morgeson, Delaney-Klinger, Mayfield, Ferrara, & Campion, 2004), there has been no comprehensive investigation exploring how incumbent judgments vary across the range of requirements commonly rated in work analysis. Enhancing our understanding of such ratings is a critical issue for practice, in part because of the significant role work analysis plays in the development of valid and legally defensible HR systems. To address this gap in the literature, we first discuss the types of inferences that are made when incumbents render judgments about the importance of various job and worker requirements as well as highlight differences among job and worker requirements that could affect incumbent judgments. Then, in a large field study, we empirically assess how systematic sources of variance and reliability of importance ratings fluctuate across five distinct types of job and worker requirements (tasks, responsibilities, knowledge, skills, and traits).

Work Analysis Ratings on Diverse Descriptors

The analysis of work is often viewed as a relatively straightforward process, when in fact it involves potentially complex inferences on the part of raters. As Sanchez and Levine (1994, p. 48) noted, "The making of job ratings can be conceptualized as an inferential decision." During typical work analysis efforts, incumbents, supervisors, and analysts are asked to make numerous inferences about their work and the work of others (Brannick, Levine, & Morgeson, 2007; Harvey & Wilson, 2000; Morgeson & Campion, 2000; Sanchez & Levine, 2001). For example, when making judgments, incumbents must first recall work-related events or behaviors, and then infer such things as (a) how often a particular task is performed, (b) the importance of certain behaviors, (c) the level of skill required in their work, or (d) what traits are important in performing the work. In this sense then, work analysis ratings involve some form of inference about the requirements of the job and/or the requirements of the worker (Morgeson & Campion, 2000; Morgeson & Dierdorff, in press). As a result, the implications of these inferences that are necessary for rendering ratings have begun to be increasingly recognized and examined in the work analysis literature (Lievens & Sanchez, 2007; Lievens, Sanchez, & DeCorte, 2004; Morgeson et al., 2004; Sanchez & Levine, 1994; Voskuijl & van Sliedregt, 2002).

Recent research has provided some indirect evidence of rating differences attributable to the type of job and worker requirement being judged during work analysis (Dierdorff & Morgeson, 2007; Dierdorff & Rubin,

2007; Dierdorff & Wilson, 2003). Although insightful, these studies have examined limited types of requirements. As such, existing research only offers a partial picture of the practical ramifications of rating differences across the full spectrum of job and worker requirements. Understanding such potential ramifications is important for at least three reasons. First, judgments in the domain of work analysis have been typically viewed as free from error until relatively recently (Harvey, 1991; Morgeson & Campion, 1997). As research has begun to indicate that work analysis judgments are subject to various systematic sources of error and inaccuracy (e.g., Morgeson et al., 2004; Van Iddekinge, Putka, Raymark, & Eidson, 2005), it becomes imperative to study the ways in which these judgments can be limited.

Second, incumbents continue to be a valuable source of work analysis information (Brannick et al., 2007), primarily because of accessibility, cost effectiveness, and direct exposure to the job itself (Sanchez & Levine, 2001). However, there is some evidence indicating that actual incumbents, as compared to professional analysts, have more difficulty in making the judgments required in work analysis (Dierdorff & Wilson, 2003; Jones et al., 2001). This evidence suggests that rating differences across various job and worker requirements may be more noticeable for incumbent raters.

Third, the popularity of newer forms of work analysis such as "competency modeling" (Lucia & Lepsinger, 1999; Schippmann, 1999) or "work profiling" (Visser, Altink, & Algera, 1997) has increased the emphasis placed on person-oriented requirements (Schippmann et al., 2000). Although such data have long been captured in more traditional work analysis techniques (Sackett & Laczo, 2003), these contemporary approaches have increased the focus on a broader set of requirements. Further, there has been some suggestion that the broader requirements rated within competency modeling may entail more complex inferences than more narrowly focused requirements such as tasks (Lievens et al., 2004).

Because so many HR activities are predicated on judgments of different job and worker requirements, understanding the extent to which incumbent ratings may vary across requirements commonly rated in work analysis becomes a key question for both practice and research. Toward this end, this study seeks to present data to help broaden our understanding of how incumbent ratings may be affected by the diverse descriptors upon which work-related judgments are made. We next discuss the various types of descriptors that are commonly rated in work analysis and emphasize two important differences among these descriptors.

Types of Job and Worker Requirement Descriptors

Broadly speaking, descriptors are simply the various features of work examined during a work analysis (Brannick et al., 2007). A number of

different descriptors have been identified in work analysis literature and they cover a wide range of the world of work (i.e., from tasks to abilities and so forth). Over 3 decades ago, Dunnette (1976) characterized these numerous descriptors as representing "two worlds of human behavioral taxonomies." More recently, Sackett and Laczo (2003) categorized descriptors commonly used in work analysis as comprising either *activities* or *attributes*, with activities referring to job-oriented descriptors (e.g., tasks and duties) and attributes referring to worker-oriented descriptors (e.g., skills and traits).

In a general sense, activity descriptors provide a focus on how work is behaviorally enacted and typically consist of tasks and responsibilities (Dierdorff & Morgeson, 2007). Tasks represent a collection of several specific elements, including an action, the object of action, and the purpose of the action (Fine & Getkate, 1995). For example, a task for an audiologist might be "fit and dispense assistive devices, such as hearing aids, to improve hearing impairments." Responsibilities are general activity statements that are aggregates of several highly related behaviors used in accomplishing major work goals (Jeanneret, Borman, Kubisiak, & Hanson, 1999). This type of descriptor is also applicable across a wide variety of jobs or occupations (Cunningham, 1996). For example, responsibilities could include items such as "analyzing data or information" or "judging the qualities of objects, services, or people."

Whereas tasks and responsibilities have an obvious behavioral focus, attribute descriptors indicate what personal requirements are needed for performing the job (Dierdorff & Morgeson, 2007). These person-oriented descriptors typically include knowledge, skill, ability, and other characteristics such as personality traits (KSAOs). Costanza, Fleishman, and Marshall-Mies (1999) define knowledge as collections of discrete but related facts and information about particular domains (e.g., biology, engineering, and technology). Skills are generally defined as the capability to perform a learned activity (e.g., communication skills and problem-solving skills), including procedures for acquiring and working with information (Mumford, Peterson, & Childs, 1999). Abilities encompass unobservable worker characteristics (e.g., inductive reasoning and perceptual speed) that define an individual's capacity to perform a wide range of activities (Fleishman, Costanza, & Marshall-Mies, 1999). Finally, other personal characteristics generally encompass personality or motivational traits presumed to affect performance (e.g., stress tolerance, Conscientiousness, and initiative).[1]

---

[1]It should be noted that our descriptions, and later operationalizations, of activity and attribute descriptors are derived from the approach used in developing the occupational information network (O*NET) content model.

Specificity and Observability of Descriptors

Although there are many distinctions that could be drawn across the diverse descriptors used in work analysis, two salient features are particularly relevant. First, there are differences in the level of specificity or detail with which each descriptor captures overall job and worker requirements. Specificity reflects the extent to which the descriptor represents a discrete unit of work (see McCormick, 1979). Descriptors of high specificity are also more likely to be linked only to select jobs, whereas descriptors of low specificity tend to be applicable to a broader range of jobs (Mumford & Peterson, 1999). For example, tasks are highly specific, represent the most molecular descriptors commonly used in work analysis (Dierdorff & Wilson, 2003), and are not typically applicable across different jobs (e.g., accountants and paramedics will have different sets of tasks associated with their jobs). Traits, on the other hand, are low in specificity because they encompass more diffuse facets of work (Morgeson & Campion, 1997) and can be used to describe numerous and quite different jobs (Sanchez & Levine, 2001).

Second, there are differences in the extent to which a particular descriptor is observable. Observability reflects the degree to which there are clear behavioral references to the work itself. For example, responsibilities are high in observability because they concern explicit collections of visible work behavior. In contrast, abilities are low in observability because they represent psychological constructs that are not directly visible and are presumed to relate causally to job performance (Fleishman et al., 1999; Morgeson & Campion, 2000).

In relation to work analysis ratings, specificity and observability can be broadly described in terms of how they operate across the activity–attribute dichotomization of descriptors. Because activity descriptors are visible and explicit work phenomena, they are considerably more specific and observable than attribute descriptors (Harvey & Wilson, 2000). For example, it is markedly easier to observe an individual perform a given task than it is to "view" aspects of their underlying ability or personality that are presumably driving the behavioral performance of the task. Instead, attribute descriptors require more complex inferences drawn from the demands of the work. Given these contrasts, work analysis researchers have argued that more abstract and complex inferences are involved when making judgments about attributes as compared to activities, which could systematically impact the ratings made on such descriptors (Harvey, 1991; Morgeson & Campion, 2000; Morgeson & Dierdorff, in press).

There are additional finer-grained distinctions that can be made regarding differing levels of specificity beyond the simple activity–attribute

dichotomization. For example, tasks and responsibilities differ in the level of specificity at which they describe work behavior (Dierdorff & Morgeson, 2007). Tasks are more molecular than responsibilities in that they describe a *single* purposeful action as opposed to a *collection* of similar purposeful actions. Thus, responsibilities are less specific than tasks because they subsume a gamut of more precise task elements. Consider the responsibility of "getting information" as an example. This responsibility could include more specific tasks ranging from "analyzing job requirements" or "identifying training needs" to "interviewing an incumbent" or "collecting statistical data."

Attribute descriptors vary even more substantially than activity descriptors in terms of specificity. For example, although skills are typically defined through their behavioral manifestations (e.g., oral communication skills indicated by proper speech), they also represent individuals' *capabilities* for behavioral acquisition (Ackerman, 1987). The demonstration of "skilled performance" is gained through expertise acquired from deliberate behavioral practice (Ericsson & Charness, 1994). Moreover, a given skill can impact behavioral performance on a range of tasks or responsibilities (Mitchell, Ruck, & Driskill, 1988). Thus, skills are lower in specificity than tasks and responsibilities because they also tap into broader sets of personal qualities relevant to job performance, even though skills are closely linked to observable actions.

Similar to skills, knowledge descriptors are more molar than tasks and responsibilities. However, differences in specificity between these two attribute descriptors are not as pronounced. For example, both knowledge and skill are developed through formal learning or the accumulation of experiences (Costanza et al., 1999) and are considered less stable than other attribute descriptors such as abilities and personality traits. Knowledge and skill are also generally applicable to the performance of certain jobs (e.g., knowledge of fine arts for curators or troubleshooting skills for aircraft mechanics). In addition, knowledge and skill are highly related in terms of how they impact particular work behavior. Mumford et al. (1999) depicted the functional interrelationship between knowledge and skills as "skills cannot be defined apart from some performance domain involving the acquisition and application of certain types of knowledge" (p. 50).

Abilities and other characteristics, such as traits, represent the most molar worker requirements (Advisory Panel for the Dictionary of Occupational Titles [APDOT], 1993). These person-oriented descriptors capture the least specific and observable characteristics of individuals that are presumed necessary for successful job performance. Both abilities and traits encompass enduring personal characteristics. In contrast to tasks and responsibilities, such descriptors are hypothetical constructs not subject to

direct observation. Unlike knowledge and skills, abilities and traits are less situational and tend to be stable over time (Fleishman et al., 1999; Mumford & Peterson, 1999). Because abilities and traits are latent psychological variables and do not always have straightforward behavioral references (Sanchez & Levine, 2001), they capture requirements at the lowest level of specificity and observability.

The preceding discussion focuses on differences in specificity and observability between various types of descriptors. It is important to recognize, however, that these distinctions may also occur within a given type of descriptor. For example, within the skills domain, certain skills (e.g., computer programming) are likely to be more specific and observable than other skills (e.g., reading comprehension). Such variability is most likely to occur when the work analysis does not utilize standardized or generic descriptor operationalizations. Yet, this potential within-descriptor variability is likely to be more pronounced for specificity differences than observability differences. For example, options to increase the observability of abilities or traits are quite limited, whereas ways to increase the specificity of these descriptor types are less constrained (e.g., rating a requirement of Conscientiousness vs. requirements of self-discipline, order, achievement orientation, and so forth). Nonetheless, different types of descriptors capture distinct facets of work, and thus, examining between-descriptor-type differences is meaningful when one considers the general shift in work analysis to using more standardized and generic descriptors applicable across a wider range of jobs or occupations (see Cunningham, 1996).

Practical Ramifications of Rating Diverse Descriptors

Because specificity and observability vary across the spectrum of descriptors, what remains to be examined is how work analysis ratings may be affected by these differences. One way to explore these rating consequences is to examine different sources of variance that contribute to work analysis ratings. That is, assessing how different sources contribute to rating variance allows us to depict the effects of using different descriptors for work analysis.

Sources of Variance in Work Analysis Ratings

Variance in work analysis ratings can stem from both unsystematic sources (random error) and from systematic sources. Typically, variance in ratings for a given job is simply attributed to random error without any further investigation (Harvey, 1991). This ignores the possibility that some of this variance may actually be due to systematic sources (Morgeson

& Campion, 1997; Morgeson, Delaney-Klinger, & Hemingway, 2005). Given the preceding discussion, one such systematic source is the type of descriptor on which incumbents are providing ratings. These effects would become evident through changes in the relative contribution of different sources of variance for each type of descriptor (e.g., tasks, skills, etc.).

The central purpose of work analysis is to scientifically gather and structure information about work (Gael, 1988; McCormick, 1976). Therefore, to meet this purpose one would hope that the majority of *systematic* variance for any work analysis rating would be directly attributable to the descriptor rather than to the person making the rating (Lievens et al., 2004). In such a case, more influence stems from rating differences across descriptors (e.g., the importance of various skills within a job) and is considered to represent "true" cross-job differences in job and worker requirements. When incumbents are performing similar work, systematic variance that is attributable to raters is generally considered idiosyncratic and can represent poor reliability (Van Iddekinge et al., 2005).

The effects of specificity and observability differences across diverse descriptors should be evident in the amount of systematic variance attributable to these sources. For example, if tasks and responsibilities are more specific and observable, variance in ratings attributable to these types of descriptors should be larger than for attribute descriptor ratings. Such a finding indicates that incumbent rating differences are linked to the requirements of a job via the specific descriptor items. On the other hand, because attribute ratings are less specific and observable, they require more subjective inferences. The consequence of these effects is to increase the variance in ratings due to idiosyncratic styles (variance due to the rater) for such descriptors.

Assuming that raters are incumbents performing similar work, a significant amount of idiosyncratic rater variance within work analysis ratings also highlights low interrater reliability. Thus, if one implication of rating less observable and specific descriptors is to increase the amount of variance attributable to raters, such influence will be evident in lower levels of interrater reliability for those descriptors (Lievens et al., 2004). Interrater reliability is a salient component of rating quality and should be important to any user of work analysis information (see Dierdorff & Wilson, 2003). For example, low reliability of ratings can have significant detrimental effects on the HR functions relying on these data, such as the misidentification of important worker characteristics necessary for job success, failure to incorporate critical activities and skills into training programs, or inappropriate choices of employment tests in personnel selection.

Method

Sample and Procedure

Archival data were used in our study and were derived from the U.S. Department of Labor's Occupational Information Network (O*NET). O*NET is a comprehensive database of occupational information and replaces the 70-year-old *Dictionary of Occupational Titles* (*DOT*; Dye & Silver, 1999). The O*NET database is organized around a theoretical content model comprising six major areas: worker characteristics, worker requirements, experience requirements, occupation requirements, occupational characteristics, and occupation-specific information (Mumford & Peterson, 1999; Peterson et al., 2001). This structure enables a focus on areas that describe important attributes and characteristics of both workers and the work itself.

Work analysis ratings for task, responsibility, knowledge, skill, and trait descriptors were examined.[2] These data came from 47,137 incumbents working in 309 different occupations. Occupations were chosen to adequately encompass a broad and representative sample of occupations that, at the time of the study, had available incumbent data. In comparison to the U.S. Department of Labor Bureau of Labor Statistics' Standard Occupational Classification system (SOC), all SOC major groupings for which incumbent O*NET data are available were represented. Table 1 displays the frequencies of the 309 occupations with respect to the SOC as well as other U.S. workforce statistics. As seen in this table, our sample is similar to the U.S. workforce at large in terms of both occupational and employment composition.

O*NET data are representative of the national labor force and were collected using a staged sampling process. First, the broad business environment in which a target occupation resides was examined to determine the types of establishments that employ occupational incumbents, the different sizes of such establishments, and how many individuals are employed in the target occupation within the U.S. labor market. Next, stratified random sampling was performed to select representative establishments for possible data collection. Third, the randomly selected establishments were contacted to verify the employment of individuals in the target occupation and to enlist participation. Fourth, individuals from the chosen establishments were randomly selected to receive various data

---

[2]Ideally, we would have included data on ability descriptors as well. Unfortunately, in the current data set comparable ability data are not available. In the O*NET sample the ability data are provided by analysts, not incumbents.

TABLE 1
Occupational Sample and Labor Market Comparisons

SOC code	SOC major group	f	Sample, %	Labor market, %	Employment, %
11-0000	Management occupations	20	6.47%	4.37%	4.44%
13-0000	Business and financial operations occupations	14	4.53%	3.87%	4.39%
15-0000	Computer and mathematical occupations	9	2.91%	2.12%	2.32%
17-0000	Architecture and engineering occupations	21	6.80%	4.49%	1.83%
19-0000	Life, physical, and social science occupations	14	4.53%	5.62%	.93%
21-0000	Community and social services occupations	5	1.62%	2.25%	1.32%
23-0000	Legal occupations	5	1.62%	1.25%	.74%
25-0000	Education, training, and library occupations	45	14.56%	7.74%	6.19%
27-0000	Arts, design, entertainment, sports, and media occupations	18	5.83%	5.24%	1.30%
29-0000	Healthcare practitioners and technical occupations	21	6.80%	6.74%	5.06%
31-0000	Healthcare support occupations	10	3.24%	2.00%	2.63%
33-0000	Protective service occupations	12	3.88%	2.75%	2.28%
35-0000	Food preparation and serving-related occupations	5	1.62%	2.37%	8.32%
37-0000	Building and grounds cleaning and maintenance occupations	3	.97%	1.37%	3.32%
39-0000	Personal care and service occupations	18	5.83%	4.24%	2.45%
41-0000	Sales and related occupations	7	2.27%	2.87%	10.64%
43-0000	Office and administrative support occupations	32	10.36%	6.99%	17.40%
45-0000	Farming, fishing, and forestry occupations	3	.97%	2.00%	.34%
47-0000	Construction and extraction occupations	18	5.83%	7.49%	5.04%
49-0000	Installation, maintenance, and repair occupations	10	3.24%	6.49%	4.04%
51-0000	Production occupations	10	3.24%	13.86%	7.74%
53-0000	Transportation and material moving occupations	9	2.91%	6.37%	7.28%

Notes. SOC = standard occupational classification, f = frequency of occupations, Sample % = percentage of occupations by total sample, Labor market % = percentage of total number of occupations in U.S. workforce, Employment % = percentage of employment in U.S. workforce; percentages derived from U.S. Bureau of Labor Statistics data.

collection questionnaires. The number of randomly selected individuals from a given establishment was based upon the proportion of incumbents in the labor market that work at such establishments.

To reduce the operational burden for respondents, items were organized into four separate questionnaires, and respondents were randomly assigned to one of the four questionnaires. However, all respondents were required to complete a task questionnaire. The random assignment of incumbents is a valuable feature of data collection as it creates independent samples across all descriptor surveys, except for tasks. To create independence for task data, we randomly selected a sample from each of the 309 occupations from which only task ratings were used (i.e., responsibility, knowledge, skill, or trait ratings were withheld). More important, this selection technique ensured sample independence and eliminated potential common source bias. Questionnaire items for all descriptors were rated using a 5-point importance scale (1 = *not important*, 2 = *somewhat important*, 3 = *important*, 4 = *very important*, and 5 = *extremely important*).

Descriptors

Tasks. Task questionnaires for each occupation were developed for O*NET by professional analysts with the goal of only including a set of tasks for a given occupation that a vast majority (80% was the goal) of incumbents could be expected to perform. In other words, because most occupations could have several hundred tasks, it was crucial to only include tasks that all incumbents would need to perform. These task lists were subsequently verified by incumbents during task list development. O*NET task questionnaires typically contain 25–30 task statements. Examples of task statements include "answer user inquiries regarding computer software or hardware operation to resolve problems"; "conduct auditing of establishments, and determine scope of investigation required"; "present and summarize cases to judges and juries"; "record patients' medical information and vital signs"; and "harvest, transplant, or pot plants." Past research has shown acceptable levels of reliability (mean intraclass correlation ICC[1,30] = .94) for O*NET tasks (Sager, Mumford, Baughman, & Childs, 1999).

Responsibilities. Responsibility requirements were collected for O*NET using a standardized questionnaire that assesses generalized work activities. These activities are organized by four categories: (a) information input, measured by 5 items; (b) mental processes, measured by 10 items; (c) work output, measured by 9 items; and (d) interacting with others, measured by 17 items (Jeanneret et al., 1999). Item examples include "getting information," "selling or influencing others," and "documenting

or recording information." Each item is presented with a brief definition to facilitate respondent comprehension. For example, the definition for the item "selling or influencing others" was "convincing others to buy merchandise/goods or to otherwise change their minds or actions." Previous research has shown acceptable levels of reliability (mean ICC[1,30] = .92) for items in this questionnaire (Childs, Peterson, & Mumford, 1999).

Knowledge. Knowledge requirements were collected for O*NET using a standardized questionnaire consisting of 33 items. These items are organized into nine categories: business and management, manufacturing and production, engineering and technology, mathematics and science, health services, arts and humanities, law and public safety, communications, and transportation (Costanza et al., 1999). Item examples include "psychology," "sales and marketing," and "personnel and human resources." All items are presented with brief definitions. For example, the definition of the item "personnel and human resources" was "knowledge of principles and procedures for personnel recruitment, selection, training, compensation and benefits, labor relations and negotiation, and personnel information systems." Prior research has shown acceptable levels of reliability (mean ICC[1,30] = .95) for items in this questionnaire (Childs et al., 1999).

Skills. Skill requirements were collected for O*NET using a standardized questionnaire containing 35 skill items. These items are organized into seven categories: content skills, process skills, social skills, complex problem-solving skills, technical skills, systems skills, and resource management skills (Mumford et al., 1999). Item examples include "judgment and decision making," "time management," and "social perceptiveness." All skill items are presented with brief definitions. For example, the definition of the item "judgment and decision making" was "considering the relative costs and benefits of potential actions to choose the most appropriate one." Research has shown acceptable levels of reliability (mean ICC[1,30] = .93) for the 35 skill items (Childs et al., 1999)

Traits. Trait requirements were collected for O*NET using a standardized questionnaire that assesses 16 work-related traits ("work styles"). These traits comprise seven dimensions: achievement orientation, social influence, interpersonal orientation, adjustment, Conscientiousness, independence, and practical intelligence (Borman, Kubisiak, & Schneider, 1999). Example items include "achievement/effort," "attention to detail," "social orientation," "dependability," and "adaptability/flexibility." Each work style is presented with a brief definition. For example, the definition of the item "adaptability/flexibility" was "job requires being open to change (positive or negative) and to considerable variety in the workplace." Previous research has shown acceptable levels of reliability (mean

ICC[1,30] = .86) for items contained in this questionnaire (Childs et al., 1999).

Occupational complexity. The study sample spanned a wide variety of occupations that could meaningfully differ in terms of their complexity. This complexity could very well influence work analysis ratings. Thus, when examining rating differences (sources of variance or reliability) across occupations, complexity could serve as a moderating influence (Dierdorff & Wilson, 2003). To examine this potential influence, we created a complexity scale using occupational-level scores derived from O*NET data describing ability requirements (Fleishman et al., 1999). More specifically, we estimated mean scores for each occupation from the 21 "cognitive ability" items contained in O*NET. These items span seven categories of cognitive ability requirements: verbal abilities, idea generation and reasoning abilities, quantitative abilities, memory, perceptual abilities, spatial abilities, and attentiveness. These ability ratings are independently derived from professional analysts. Previous research has shown acceptable levels of reliability (mean ICC[1,30] = .93) for O*NET ability items (Childs et al., 1999). Coefficient alpha for the complexity scale was .94.

Analytic Strategy

Variance component (VC) analysis was used to simultaneously estimate the contribution of systematic sources of variability in incumbent ratings. VC analysis allows for the partitioning of observed variance in a set of observations (Searle, Casella, & McCulloch, 1992). Estimated VCs can also be used to compute the proportion of systematic variance due to factors specified by the researcher. Because the samples were independent for each type of descriptor within an occupation, separate VC analyses were conducted for each occupation within each type of descriptor (tasks, responsibilities, knowledge, skills, and traits).

For each of the 309 occupations, VC models partitioned variance into two systematic sources: variance due to the rater (incumbent) and variance due to the item (descriptor). Thus, the total variance in ratings of a given descriptor type (e.g., skills) was decomposed into three sources: (a) systematic rater variance, (b) systematic item variance, and (c) residual variance. To reiterate, a significantly large VC attributable to raters indicates that ratings are heavily influenced by idiosyncratic rating styles, versus the variability in the descriptor being judged. A significantly large VC attributable to items indicates discriminant validity of items across a given job or worker requirement (i.e., "true" variance). Variances due to items were not estimated for task ratings because tasks are occupation specific by nature and thus vary across occupations so as to preclude

such estimation. All VC models were fit using the MIXED procedure in SAS 9.1® (Littell, Milliken, Stroup, & Wolfinger, 1996). Each factor was considered random in order to generate the appropriate VCs (Shavelson & Webb, 1991). Restricted maximum likelihood (REML) was used to estimate VCs as research has suggested that this estimation procedure produces more accurate estimates (DeShon, 1995; Searle et al., 1992).

In addition to the VC analyses, we used meta-analytic procedures (Hunter & Schmidt, 1990) to estimate levels of interrater reliability for each descriptor type. Occupational mean profile correlations were meta-analyzed so as to estimate cumulative interrater reliability across the 309 occupations. Prior to meta-analysis, profile correlations were first computed as Pearson product–moment correlations between an incumbent's profile of ratings and the profile of mean ratings for the entire occupational sample excluding the incumbent. Thus, the lower the magnitude of an individual's profile correlation, the less his or her ratings resemble those of the occupational sample (Ballentine, Cunningham, & Wimpee, 1992; Dierdorff, Wilson, & Carter, 2004; Earles, Driskill, & Dittmar, 1996). Next, mean profile correlations were computed for each occupation on each descriptor, which meant that there were five mean profile correlations for every occupation (i.e., one for tasks, one for responsibilities, one for knowledge, one for skills, and one for traits). These occupational mean profile correlations represent the average level of interrater reliability on each type of descriptor for a given occupation. The occupational mean profile correlations were used as input for the study's meta-analyses. Similar to previous meta-analyses of interrater reliability of work analysis ratings (e.g., Dierdorff & Wilson, 2003), the reliability estimates were corrected only for sampling error. The study's meta-analyses produced sample-size weighted mean estimates of interrater reliability for the each of the five descriptors. These estimates can be viewed as representing the overall interrater reliability across occupations that one can expect when different types of descriptors are under consideration.

Results

Separate VC analyses were conducted for each occupation within each of the five descriptors (tasks, responsibilities, knowledge, skills, and traits). All individual VCs (rater, item, residual) within each occupation's estimated model for all five types of descriptors were significant ($p < .01$). Table 2 displays the VC results collapsed across occupations but delineated by descriptor. The VC values and the associated percentages of variance explained shown in Table 2 are mean values calculated across occupations. As previously discussed, variance due to the rater is generally undesirable and reflects idiosyncratic variance, whereas variance due to the item is

TABLE 2
Variance Components by Work Descriptor

Work descriptor	Variance source	VC	% variance explained
Task	Item	–	–
	Rater	.41	11.61%
	Residual	3.12	
Responsibility	Item	.68	35.42%
	Rater	.31	16.15%
	Residual	.93	
Knowledge	Item	.75	42.13%
	Rater	.26	14.61%
	Residual	.77	
Skill	Item	.72	39.13%
	Rater	.34	18.48%
	Residual	.78	
Trait	Item	.14	15.23%
	Rater	.32	34.78%
	Residual	.46	

Note. VC = variance component; all VC estimates are averages across the 309 occupations within each type of work descriptor.

desirable in that it displays discriminability. Important to assessing the impact of specificity and observability, proportionally less variance due to the item and more due to the rater is indicative of effects attributable to a particular type of descriptor.

The results in Table 2 show that at the ends of the specificity continuum, incumbents displayed the least idiosyncratic variability when rating tasks (11.61%) and the most when rating traits (34.78%). Proportions of variance due to rater were relatively comparable across responsibility, knowledge, and skill descriptors (ranging from 14.61% to 18.48%). With regard to variance due to items, ratings on the most molar and least observable descriptor (i.e., traits) by far had the smallest proportion (15.23%). Proportions of variance due to items were again relatively comparable across responsibility, knowledge, and skill descriptors (ranging from 35.42% to 42.13%). Overall, findings indicate that rating consequences due to descriptors are greatest at the specificity and observability extremes (tasks and traits), but rather similar across responsibility, knowledge, and skill descriptors.

We also explored whether the proportional differences of variance due to rater across descriptors were related to differences in the complexity of occupations in which raters worked. That is, we assessed whether occupational complexity was associated with increases or decreases in

TABLE 3
Meta-Analysis of Interrater Reliability

Work descriptor	R_{wt}	n	k	SD_{wt}	95% CI
Task	.80	9,333	309	.071	.79–.81
Responsibility	.65	9,713	309	.085	.64–.66
Knowledge	.70	9,506	309	.077	.69–.71
Skill	.69	9,157	309	.107	.68–.70
Trait	.45	9,428	309	.096	.44–.46

Note. R_{wt} = sample-size weighted mean estimate of interrater reliability; CI = confidence interval.

idiosyncratic variance. The proportion of variance due to rater was significantly and inversely related to occupational complexity for all descriptors ($p < .01$). The correlation between proportions of idiosyncratic variance and complexity was −.50 for responsibility ratings, −.39 for skill ratings, −.43 for knowledge ratings, and −.29 for trait ratings. These results suggest that ratings from more complex occupations contain less variance due to the rater.

Table 3 displays the results of the meta-analyses examining the interrater reliabilities of ratings. The table provides the sample-size weighted mean estimates of interrater reliability, observed standard deviations of these estimates, and 95% confidence intervals. Mean interrater reliability estimates were highest for the task ratings (.80) and lowest for trait ratings (.45). Mean interrater reliabilities for the other descriptors were relatively similar in magnitude, ranging from .65 to .70. The mean interrater reliability for ratings of skills and responsibilities did not significantly differ, as evidenced by their overlapping 95% confidence intervals. In general, findings from the meta-analyses mirror those from the VC analyses, in that differences in the reliability of ratings were most pronounced when comparing task to trait descriptors.

Discussion

The notion of an "inferential leap" has been forwarded to point out that work analysis ratings involve a number of judgments that vary in the amount of inference required (Brannick et al., 2007; Morgeson & Campion, 2000; Sanchez & Levine, 2001). Harvey (1991) reflected this notion by concluding that "it is a long inferential leap from a description of observable job tasks to a listing of general ability and trait requirements." (p. 78). In a broad sense, our findings offer indirect empirical evidence of this inferential leap for incumbent ratings across different kinds of job and worker requirements. Drawing on work analysis literature, we proposed

that ratings would systematically vary due to the descriptors that were being judged. We further proposed that these effects could be described vis-à-vis the specificity and observability with which descriptors depict job and worker requirements. Taken collectively, our findings from a large-scale field study suggest that the types of descriptors on which incumbents must draw inferences can have noticeable consequences for work analysis ratings.

Specifically, the consequences for incumbent ratings become apparent when one examines the changing magnitudes of rater, item, and residual VCs, as well as varying levels of interrater reliability, across the five distinct descriptors included in our analyses. Here, as specificity and observability fluctuated across descriptors so did the estimates of interrater reliability and the amounts of variance attributable to different sources. However, these differences were greatest for incumbent raters at the specificity and observability extremities (i.e., task ratings compared to trait ratings). For example, the average percentage of variance attributable to raters (idiosyncratic variance) was more than double the magnitude attributable to items (34.78% vs. 15.23%) for incumbent ratings of trait descriptors. In addition, trait ratings displayed the lowest interrater reliability and were the only descriptor for which more variance was due to the rater than to the item.

Another interesting finding comes from a closer inspection of the results for descriptors between the task–trait descriptor extremities. Results from both variance decomposition and meta-analysis indicate that rating consequences are quite comparable across responsibility, knowledge, and skill ratings. This comparability can be seen in the similarity of interrater reliabilities and the proportions of variance attributable to different sources across these descriptor types. From the viewpoint of between-descriptor-type differences, it may be that responsibilities, knowledge, and skills describe work at a somewhat middle ground with regard to specificity, as they can apply to multiple jobs or occupations. However, these types of descriptors certainly differ to a greater extent with respect to observability, as responsibilities describe actual work behavior whereas knowledge and skills describe person-oriented requirements. That specificity appears more central to rating consequences is also congruent with other work analysis research showing differential relationships at varying levels of descriptor specificity (e.g., Dierdorff & Morgeson, 2007). Yet, it is important to reiterate that our findings focus on broad differences between types of descriptors. Certainly, specificity could vary within a particular kind of descriptor as well, a point we will return to shortly.

Our findings also indicate that occupational complexity is associated with proportions of idiosyncratic variance in work analysis ratings. Here, results show that ratings rendered by incumbents working in more

complex occupations displayed lower proportions of idiosyncratic variance. Of additional note is that this relationship appears most pronounced for ratings of responsibilities ($r = -.51$) and least pronounced for trait ratings ($r = -.29$). Thus, the results suggest that trait ratings not only have the largest proportions of undesirable idiosyncratic variance but also that these proportions are not as impacted by the complexity of occupational roles. This is perhaps not surprising, as traits are the most general type of descriptor. An illustration of these differences can be made by comparing two occupations in our sample that differ markedly in complexity. Among the most complex occupations in our sample was "chemical engineers," where proportions of idiosyncratic variance were 13.2% for responsibility ratings and 28.6% for trait ratings. In comparison, among the least complex sample occupations was "cooks, fast food" where the proportions were 33.9% for responsibility ratings and 33.4% for trait ratings.

Implications for Work Analysis Practice

Some work analysis experts question whether it is even reasonable to expect incumbents to be capable of making high-quality inferences about various job and worker requirements, especially when such judgments are for worker-oriented attributes (Harvey & Wilson, 2000). This concern flows from the fact that in practice it is a rather rare circumstance to encounter an incumbent describing his or her job in terms of its most important knowledge, skills, traits, and so forth. Most individuals tend not to think of their jobs in the complex dimensions utilized by most analysts. Although this research cannot offer a definitive conclusion to the question of incumbent capability, it does offer some compelling evidence to suggest that using descriptors low in specificity and observability will systematically impact work analysis ratings. Taken collectively, these findings make a case against the capability of incumbents to reliably rate highly abstract trait descriptors. On the other hand, our findings also show that incumbents seem quite capable of providing reliable and discriminating ratings on other attribute descriptors, such as skills and knowledge. Thus, our findings support concerns about some attribute descriptors (e.g., traits) but not others (e.g., knowledge and skills).

Of course, in work analysis practice the choices of descriptor and source must be predicated upon the intended use of the work analysis information (Morgeson & Dierdorff, in press; Sackett & Laczo, 2003). For example, when the intended use of work analysis information is to inform the choice of various selection instruments, identifying the most important skills, knowledge, and traits is essential, whereas in developing training programs, tasks, responsibilities, and skills are frequently the

primary focus. One important implication from our study is that the extent to which the chosen respondents agree upon the most critical or important requirements of target jobs then becomes an issue of decision-making quality. A lack of reliability in judgments of job and worker requirements, or large amounts of rating contamination due idiosyncratic rater factors, will certainly have adverse effects on the HR functions relying on these data. Furthermore, professional guidelines require that specific consideration and caveats are warranted when a lack of consensus exists regarding work analysis information (e.g., *Principles for the Validation and Use of Personnel Selection Procedures,* Society for Industrial and Organizational Psychology, 2003). These findings suggest that HR practices relying heavily on trait descriptors derived from incumbents could lead to lower-quality decisions based upon such information.

Interestingly, the heavier emphasis on trait descriptors is one of the defining characteristics of competency modeling (Schippmann et al., 2000). Our findings for these abstract features of work suggest that if incumbents are chosen as respondents, the quality of work analysis data based on these ratings is likely to suffer. For the work analysis practitioner, this creates a decisional quandary around whom to use as source of work information. One solution may be the use of professional analysts when abstract descriptor ratings are necessary, as some evidence indicates analysts provide more reliable ratings (Dierdorff & Wilson, 2003). At the very least, analysts should be more accustomed than job incumbents to conceptualizing work with regard to the full spectrum of activity and attribute requirements.

A second potential solution involves making the rating situation less complex for incumbents by breaking down abstract judgments into more manageable components (Cornelius & Lyness, 1980). The Fleishman Job Analysis Survey (F-JAS) technique for deriving ratings on abstract ability descriptors by anchoring scales with task requirements is one such approach (Fleishman, 1992; Fleishman & Reilly, 1992). Another example is outlined by Lievens et al. (2004) and blends task information with competency modeling, showing improved reliability and discriminability of ratings.

A third alternative is to provide rater training for work analysis respondents (Gael, 1988). Recent empirical evidence suggests that providing raters with frame-of-reference (FOR) training is a viable intervention capable of improving ratings on abstract descriptors. For example, Lievens and Sanchez (2007) found FOR training to increase discriminant validity, interrater reliability, and accuracy of competency ratings. In addition, Aguinis, Mazurkiewics, and Heggestad (2009) found that FOR training lowered mean ratings of trait descriptors and reduced the average

correlation between self-ratings and job ratings of trait descriptors (i.e., decreasing idiosyncratic variance).

A final practical implication of our findings relates to how descriptor items are constructed for use in work analysis projects. Our results clearly indicate that specificity plays an important role in the quality of work analysis information gleaned from incumbent raters. Such a finding suggests that when designing descriptor items there may be considerable practical benefits for emphasizing specificity, even when the descriptors themselves are more molar in nature. In fact, based on our results these practical benefits are likely to be most salient when designing abstract descriptors, such as traits or abilities, which have the most undesirable rating consequences (i.e., lower reliability, higher proportions of idiosyncratic variance). One straightforward way to infuse specificity into descriptors would be to present to incumbent raters thorough definitions of descriptor items that include distinct examples from the workplace.

Implications for Future Research

Our description of how work analysis ratings fluctuate across diverse descriptors of differing levels of specificity and observability has a number of implications for future research. First, work analysis research is often derided as somewhat atheoretical, more technique than substantive research area. One potential way to make future work analysis research more theory based is to focus on work judgments as inferential decisions and then study the range of ways in which the inferential judgments vary as a function of a range of different social, cognitive, and individual factors (Morgeson & Campion, 1997).

Second, as this study shows, judgments of different descriptors do in fact vary for reasons other than the job itself. Future research could be conducted to better understand the underlying reasons why there is variance in these judgments. Although recent research has successfully drawn from role theory to understand some of the individual and contextual factors that explain variance in job and worker requirements (Dierdorff & Morgeson, 2007; Morgeson et al., 2005), clearly much more needs to be done. Such pressing needs can be seen in recent findings showing that the majority of variance in KSAO ratings (69–80%) is due to sources other than job differences (Van Iddekinge et al., 2005). Important to note is that research attempting to account for this non-job variance requires data that are cross-role (i.e., multiple, differentiated jobs), as in this study, in order to isolate influential factors beyond the job itself that affect work analysis ratings.

Third, this research focused on exploring differences between types of descriptors. As previously discussed, there is likely to be variance

within a particular type of descriptor in addition to between-descriptor variance. To the extent that analyses collapse across specific items in a descriptor domain, this within-domain heterogeneity is ignored. O*NET descriptors are specifically designed to be generic in order to enable applicability across a wide range of occupations (Mumford & Peterson, 1999). Thus, within-descriptor-type heterogeneity may not be as pronounced for standardized instruments, such as O*NET or the, Position Analysis Questionnaire, when compared to other "home-grown" work analysis efforts or other unstandardized work analysis instruments used in practice. Future research could explore how both within- and between-descriptor differences in specificity and observability impact the magnitude of different sources of ratings variance. Perhaps some descriptor types may allow for more within-descriptor variability than others, which could result in differential rating consequences (e.g., reliability).

Fourth, our examinations more broadly emphasized the specificity and observability differences between various types of descriptors. Yet, there are other more precise features of descriptors that we did not investigate but could potentially influence work analysis ratings. The presence or absence, level of detail, and wording of descriptor definitions are examples of factors that could impact subsequent ratings. Morgeson and colleagues (2004) provided evidence showing that even simple manipulations of descriptor wording (e.g., adding the phrase "ability to" to task statements) could lead to significant differences in work analysis ratings. Thus, future research could examine the extent to which idiosyncratic variance and decreased reliability are due to the adequacy or interpretability of descriptor definitions as opposed to the descriptors themselves. The use of rating scale anchors, the overall format of work analysis instruments (e.g., length, ordering of descriptors types, etc.), as well as the kinds of representative work examples provided to raters are further examples of features that could be isolated and examined by future work analysis research.

Fifth, this research focused explicitly on incumbent ratings. As incumbents are perhaps the most commonly used source of work analysis information, such a focus makes practical sense. Yet, research is needed to see if the same rating consequences manifest with other sources of work analysis information (e.g., supervisors or analysts). It may be that these other sources are better able to make the full range of inferences often required in work analysis.

Finally, this study explored judgments of importance. This is but one kind of information that can be collected when gathering work analysis data. Other common judgments include frequency of performance, difficulty, criticality, and the extent to which a particular attribute is needed at entry to the job. It is possible that reliability and VCs vary as the nature

of the judgment being made changes. That is, rating scales may require different "inferential lenses" (Dierdorff & Rubin, 2007) through which to make judgments about various job and worker requirements. For example, some research has shown that importance ratings have higher reliabilities than other commonly used scales (Dierdorff & Wilson, 2003).

Strengths and Limitations

This study has several strengths that are worthy of mention. The number of incumbents providing work analysis ratings is rather substantial, including over 47,000 individuals. These data are derived from a well-established, standardized database designed to be representative of the national labor force. In addition, sampled incumbents come from 309 different occupations. The lack of ability descriptors notwithstanding, this study also includes the most common descriptors typically captured in work analysis efforts. These descriptors span activities and attributes, a range that allows for a more meaningful depiction of how work analysis ratings vary across types of descriptors. Finally, fluctuations in ratings are presented with results from multiple methods that include VC analysis and meta-analysis of reliability.

Of course, there are important boundary conditions associated with this study. First, descriptors and their associated items were accompanied by specific definitions to aid respondents. This is not always common practice, but rather respondents to work analysis surveys are often required to rate items that are simply listed as responsibility, knowledge, skill, or trait statements without the provision of examples or descriptions (Schippmann et al., 2000). Presenting incumbents with such limited information would likely increase idiosyncratic variance and decrease reliability compared to the current ratings.

Second, the O*NET surveys on which ratings were captured used single-item measures for each descriptor (e.g., a single item assessing the skill of "negotiation"). Many of these items could be described as multidimensional constructs and thereby assessed with multiple indicators. Using single-item direct estimation to collect information on various descriptors has been shown to lower reliability (Butler & Harvey, 1988; Cornelius & Lyness, 1980; Sanchez & Levine, 1994). In these findings, reliabilities were relatively high for most descriptors despite this direct estimation, except perhaps for trait ratings.

Third, the incumbent respondents in this study were derived from an existing database of work analysis ratings. Therefore, our results are contingent upon the extent to which these ratings are of sufficient quality. We believe the data are indeed of high quality, as they are derived from the nationally representative O*NET database and surveyed using a thorough

and standardized methodology. Yet, many other sources, such as analysts or training experts, are often used to provide work analysis data.

Fourth, although based on arguably the most expansive theoretical effort to systematically describe the world of work, the descriptors designed for O*NET are certainly not the only way to conceptualize responsibility, knowledge, skill, and trait requirements. Similar to other standardized work analysis instruments (e.g., Position Analysis Questionnaire), the design of O*NET descriptors emphasizes a "common language" by using sets of cross-job variables to capture job and worker requirements. Thus, each type of descriptor contains a rather parsimonious and generic set of items (e.g., 35 items for the O*NET skill domain). The extent to which work analysis efforts typical to HR practice diverge from this common language approach (i.e., generating job- or firm-specific descriptors) is largely unknown.

Taken collectively, the limitations presented above could impact the generalizability of our study's results. However, we feel that there are several reasons to believe that these findings will generalize to some degree. First, there is evidence to suggest that there is considerable redundancy across different work analysis rating scales. For example, several studies have shown importance ratings to demonstrate moderate-to-large overlap with other rating scales, such as time spent, frequency, and difficulty scales (Friedman, 1990; Sanchez & Fraser, 1992; Sanchez & Levine, 1989). This suggests that the importance ratings measured in this study share a number of similarities with other commonly used work analysis rating scales. Second, the underlying social and cognitive psychological process that make it difficult to render certain work analysis judgments are also likely to occur with nonincumbent sources (see Morgeson & Campion, 1997). Third, the specificity and observability distinctions we examined allow for extrapolation to other types of descriptors. Thus, to the extent that any descriptor is low in specificity and observability, the inferences required for making work analysis ratings are likely to be more complex. Despite these reasons for confidence in the generalizability of our findings, a definitive statement clearly requires further empirical research.

Finally, one possible reaction to our findings is that they merely reflect a "truism" in work analysis. That is, ratings of general and nonobservable job and worker requirements involve more difficult inferences than ratings of specific and observable requirements. Although we agree that this is a widely held belief among work analysis researchers (see Harvey, 1991), there is relatively little empirical data actually supporting this contention. The empirical data that have been offered come from recent research with limited scope in terms of the focal descriptors examined (e.g., Dierdorff & Wilson, 2003; Morgeson et al., 2004; Van Iddekinge et al., 2005). In

addition, it is unclear whether this belief among work analysis scholars is indeed similarly held by practitioners. Such disconnect between research and practice is well documented in other areas of HR management (Rynes, Colbert, & Brown, 2002; Rynes, Giluk, & Brown, 2007). We feel it is important to empirically establish the kind of work analysis issues highlighted earlier.

Conclusion

Although many have speculated about the outcomes of rating diverse descriptors used in work analysis, there has been no systematic investigation of this phenomenon. We sought to address this gap by first describing between-descriptor differences in terms of specificity and observability. We then explored differences in incumbent ratings across five types of descriptors (task, responsibility, knowledge, skill, and trait). Our findings that variance due to idiosyncratic effects increases and that interrater reliability decreases when ratings are made on molecular tasks compared to molar traits have important implications for work analysis practice, including increasing the quality of work analysis data. Yet, the need clearly remains for additional research that further explicates the various effects diverse descriptors may have on the inferences required of raters during work analysis.

REFERENCES

Ackerman PL. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3–27.

Advisory Panel for the Dictionary of Occupational Titles (APDOT). (1993). *The new DOT: A database of occupational titles for the twenty-first century* (final report). Washington, DC: U.S. Employment Service, U.S. Department of Labor Employment and Training Administration.

Aguinis H, Mazurkiewicz MD, Heggestad ED. (2009). Using web-based frame-of-reference training to decrease biases in personality-based job analysis: An experimental field study. PERSONNEL PSYCHOLOGY, *62*, 405–438.

Ballentine RD, Cunningham JW, Wimpee WE. (1992). Air force enlisted job clusters: An exploration in numerical job classification. *Military Psychology, 4*, 87–102.

Borman WC, Kubisiak UC, Schneider RJ. (1999). Work styles. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 213–226). Washington, DC: American Psychological Association.

Brannick MT, Levine EL, Morgeson FP. (2007). *Job analysis: Methods, research, and applications for human resource management* (2nd ed.). Thousand Oaks, CA: Sage.

Butler SK, Harvey RJ. (1988). A comparison of holistic versus decomposed rating of position analysis questionnaire work dimensions. PERSONNEL PSYCHOLOGY, *41*, 761–771.

Childs RA, Peterson NG, Mumford MD. (1999). Occupational descriptor covariates: Potential sources of variance in O*NET ratings. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 237–246). Washington, DC: American Psychological Association.

Cornelius ET, Lyness KS. (1980). A comparison of holistic and decomposed judgment strategies in job analysis by job incumbents. *Journal of Applied Psychology, 65*, 155–163.

Costanza DP, Fleishman EA, Marshall-Mies JC. (1999). Knowledges. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 71–90). Washington, DC: American Psychological Association.

Cunningham JW. (1996). Generic descriptors: A likely direction in occupational analysis. *Military Psychology, 8*, 247–262.

DeShon RP. (1995, May). *Restricted maximum likelihood estimation of variance components in generalizability theory: Overcoming balanced data requirements*. Paper presented at the 11th Annual Conference of the Society of Industrial and Organizational Psychology. Orlando, FL.

Dierdorff EC, Morgeson FP. (2007). Consensus in work role requirements: The influence of discrete occupational context on role expectations. *Journal of Applied Psychology, 92*, 1228–1241.

Dierdorff EC, Rubin RS. (2007). Carelessness and discriminability in work role requirement judgments: Influences of role ambiguity and cognitive complexity. Personnel Psychology, *60*, 597–625.

Dierdorff EC, Wilson MA. (2003). A meta-analysis of job analysis reliability. *Journal of Applied Psychology, 88*, 635–646.

Dierdorff EC, Wilson MA, Carter L. (2004, April). *Effects of tenure on the quality of incumbent job analysis ratings*. Paper presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Dunnette MD. (1976). *Handbook of industrial and organizational psychology*. Chicago, IL: Rand-McNally.

Dye D, Silver M. (1999). The origins of O*NET. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 9–19). Washington, DC: American Psychological Association.

Earles JA, Driskill WE, Dittmar MJ. (1996). Methodology for identifying abilities for job classification: An application of job analysis. *Military Psychology, 8*, 179–193.

Ericsson KA, Charness N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725–747.

Fine SA, Getkate M. (1995). *Benchmark tasks for job analysis: A guide for functional job analysis (FJA) scales*. Mahwah, NJ: Erlbaum.

Fleishman EA. (1992). *The Fleishman-job analysis survey (F-JAS)*. Palo Alto, CA: Consulting Psychologist Press.

Fleishman EA, Costanza DP, Marshall-Mies J. (1999). Abilities. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 175–196). Washington, DC: American Psychological Association.

Fleishman EA, Reilly ME. (1992). *Handbook of human abilities: Definitions, measurements, and job task requirements*. Palo Alto, CA: Consulting Psychologist Press.

Friedman L. (1990). Degree of redundancy between time, importance, and frequency task ratings. *Journal of Applied Psychology, 75*, 748–752.

Gael S. (1988). *Job analysis handbook for business*, *industry*, *and government*. New York: John Wiley & Sons.

Goldstein IL, Zedeck S, Schneider B. (1993). An exploration of the job analysis-content validity process. In Schmitt N, Borman WC (Eds.), *Personnel selection in organizations*. San Francisco: Jossey-Bass

Harvey RJ. (1991). Job analysis. In Dunnette MD, Hough LM (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 71–164). Palo Alto, CA: Consulting Psychologists Press.

Harvey RJ, Wilson MA. (2000). Yes Virginia, there is an objective reality in job analysis. *Journal of Organizational Behavior, 21*, 829–854.

Hunter JE, Schmidt FL. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Beverly Hills, CA: Sage.

Jeanneret PR, Borman WC, Kubisiak UC, Hanson MA. (1999). Generalized work activities. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O∗NET* (pp. 49–69). Washington, DC: American Psychological Association.

Jones RG, Sanchez JI, Parameswaran G, Phelps J, Shoptaugh C, Williams M, et al. (2001). Selection or training? A two-fold test of the validity of job analytic-ratings of trainability. *Journal of Business and Psychology, 15*, 363–389.

Lievens F, Sanchez JI. (2007). Can training improve the quality of inferences made by raters in competency modeling? A quasi-experiment. *Journal of Applied Psychology, 92*, 812–819.

Lievens F, Sanchez JI, De Corte W. (2004). Easing the inferential leap in competency modeling: The effects of task-related information and subject matter expertise. PERSONNEL PSYCHOLOGY, *57*, 881–904.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.

Lucia A, Lepsinger R. (1999). *The art and science of competency models: Pinpointing critical success factors in organizations*. San Francisco: Jossey-Bass.

McCormick EJ. (1976). Job and task analysis. In Dunnette MD (Ed.), *Handbook of industrial and organizational psychology* (pp. 651–696). Chicago: Rand-McNally.

McCormick EJ. (1979). *Job analysis: Methods and applications*. New York: AMACOM.

Mitchell JL, Ruck HW, Driskill W. (1988). Task-based training program development. In Gael S (Ed.), *Job analysis handbook for business*, *industry*, *and government* (Vol. 1, pp. 205–215) New York: John Wiley & Sons.

Morgeson FP, Campion MA. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology, 82*, 627–655.

Morgeson FP, Campion MA. (2000). Accuracy in job analysis: Toward an inference-based model. *Journal of Organizational Behavior, 21*, 819–827.

Morgeson FP, Delaney-Klinger KA, Hemingway MA. (2005). The importance of job autonomy, cognitive ability, and job-related skill for predicting role breadth and job performance. *Journal of Applied Psychology, 90*, 399–406.

Morgeson FP, Delaney-Klinger KA, Mayfield MS, Ferrara P, Campion MA. (2004). Self-presentation processes in job analysis: A field experiment investigating inflation in abilities, tasks, and competencies. *Journal of Applied Psychology, 89*, 674–686.

Morgeson FP, Dierdorff EC. (in press). Job and work analysis: From technique to theory. In Zedeck S (Ed.), *APA handbook of industrial and organizational psychology*. Washington, DC: American Psychological Association.

Mumford MD, Peterson NG. (1999). The O*NET content model: Structural considerations in designing jobs. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 21–30). Washington, DC: American Psychological Association.

Mumford MD, Peterson NG, Childs RA. (1999). Basic and cross-functional skills. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 49–69). Washington, DC: American Psychological Association.

Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA, Campion MA, et al. (2001). Understanding work using the occupational information network (O*NET): Implications for practice and research. PERSONNEL PSYCHOLOGY, *54*, 451–492.

Rynes SL, Colbert AE, Brown KG. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*, 149–174.

Rynes SL, Giluk TL, Brown KG. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal, 50*, 987–1008.

Sackett PR, Laczo RM. (2003). Job and work analysis. In Borman WC, Ilgen DR, Klimoski RJ (Eds.), *Comprehensive handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 21–37). New York: John Wiley & Sons.

Sager CE, Mumford MD, Baughman WA, Childs RA. (1999). Occupation-specific descriptors: Approaches, procedures, and findings. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 227–236). Washington, DC: American Psychological Association.

Sanchez JI, Fraser SL. (1992). On the choice of scales for task analysis. *Journal of Applied Psychology, 77*, 545–553.

Sanchez JI, Levine EL. (1989). Determining the important tasks within jobs: A policy-capturing approach. *Journal of Applied Psychology, 74*, 336–342.

Sanchez JI, Levine EL. (1994). The impact of rater's cognition on judgment accuracy: An extension into the job analysis domain. *Journal of Business and Psychology, 9*, 47–57.

Sanchez JI, Levine EL. (2001). The analysis of work in the 20th and 21st centuries. In Anderson N, Ones DS, Sinangil HK, Viswesvaran C (Eds.), *International handbook of work and organizational psychology* (pp. 71–89). Thousand Oaks, CA: Sage.

Schippmann JS. (1999). *Strategic job modeling: Working at the core of integrated human resources*. Mahwah, NJ: Erlbaum.

Schippmann JS, Ash RA, Battista M, Carr L, Eyde LD, Hesketh B, et al. (2000). The practice of competency modeling. PERSONNEL PSYCHOLOGY, *53*, 703–740.

Searle SR, Casella G, McCulloch CE. (1992). *Variance components*. New York: Wiley & Sons.

Shavelson RJ, Webb NM. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Van Iddekinge CH, Putka DJ, Raymark PH, Eidson CE. (2005). Modeling error variance in job specification ratings: The influence of rater, job, and organization-level factors. *Journal of Applied Psychology, 90*, 323–334.

Visser CF, Altink WMM, Algera JA. (1997). From job analysis to work profiling. Do traditional procedures still apply? In Anderson N, Herriot P (Eds.), *International handbook of selection and assessment* (pp. 441–454). New York: Wiley & Sons.

Voskuijl OF, van Sliedregt T. (2002). Determinants of interrater reliability of job analysis: A meta-analysis. *European Journal of Psychological Assessment, 18*, 52–62.