Call Me Old Fashioned –
Is My Job Analysis Accurate or Not?

Shanan Gwaltney Gibson

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Psychology

Robert J. Harvey, Ph.D. (chair)
Kevin D. Carlson, Ph.D.
John J. Donovan, Ph.D.
Roseanne J. Foti, Ph.D.
Morell E. Mullins, Jr., Ph.D.

Spring 2001
Blacksburg, VA

Keywords: Holistic Job Analysis, Decomposed Job Analysis, Accuracy, Interrater Agreement

Call Me Old Fashioned –
Is My Job Analysis Accurate or Not?

Shanan Gwaltney Gibson

(ABSTRACT)

As a process designed to collect information about jobs, job analysis is one of the most fundamental aspects of personnel psychology. It forms the foundation upon which almost every other human resource management component is built, including selection, compensation, performance appraisal, and training program development. Despite the considerable evidence of human fallibility in other judgment processes, many have followed the implicit assumption that job analysis information is accurate without actually examining this proposition. This study considers two potential sources of job analysis rating inaccuracy – the source of the ratings and the type of instrument utilized to collect ratings. By utilizing less job-familiar job analysis raters and shorter, more holistic job analysis instruments, industrial-organizational psychologists have attempted to attenuate the time and costs associated with the job analysis process; however, findings regarding the reliability and accuracy of such practices are questionable. Hypotheses tested in the current study indicated that decomposed measures of job behavior converged to a greater degree with an external job analysis than did holistic measures. Interrater agreements for all types of raters and across all types of instruments are concluded to be inadequate. Potential explanations from the cognitive and social psychological domains for these findings are conjectured and directions for future research are noted.

TABLE OF CONTENTS

Introduction	1
What is Job Analysis?	1
Observables vs. Inferred Characteristics	2
Job Analysis Philosophies and Instruments	3
Relativity of Rating Scales	15
Current Issues in Job Analysis	15
Cognitive Decision Making	20
Current Study	24
Conceptualizations of Job Analysis Accuracy	25
Potential Sources of Job Analysis Inaccuracy	26
Hypotheses Related to Holistic vs. Decomposed Job Analysis Measures	28
Hypotheses Related to Source of Job Analysis Ratings	31
Hypotheses Related to the Issue of Complexity	34
Hypotheses Regarding the Combination of Rater and Instrument Type	35
Methods	36
Results	40
Discussion	48
References	56
Appendix A – Job Description for “Secretary”	65
Appendix B – O*Net GWA Questionnaire	67
Appendix C – CMQ Questionnaire	68
Appendix D – Matched Dimensions for the CMQ and the O*Net	69
Appendix E – Differences in Total Mean Score by Rater Type	75
Appendix F – Correlations by Rater Type for Each Dimension	77
Appendix G – Mean Dimension Scores for Scales by Rater Type	81
Appendix H – Convergence with Researcher’s Ratings	83
Vita – Shanan Gwaltney Gibson	88

Call Me Old-Fashioned –
Is My Job Analysis Accurate or Not?

Job analysis is one of the most fundamental ingredients of personnel psychology. Although it has been called the "Rodney Dangerfield" of industrial-organizational psychology (Cunningham, 1989 in Harvey, 1991), job analysis is widely recognized as an essential foundation for numerous personnel functions (including selection, compensation, performance appraisal, and training program development). During the past three decades significant progress has been made with regard to job analysis; standardized job analysis questionnaires that are applicable to numerous positions are readily available, movement toward instruments that provide cross-job relative ratings and allow comparison of task-dissimilar positions has occurred, and efforts to make the job analysis process less intrusive and time consuming have been made. However, with these improvements have come increased debate and controversy, as well as attention that highlights the numerous gaps in our job analysis research base.

What is Job Analysis?

Even after many years of analyzing jobs, consensus can not be reached over what actually constitutes a job analysis – is a job analysis only about collecting data in order to describe the observable characteristics of a position (such as work behaviors and activities), or does job analysis rightly include inferences regarding the abilities and personal attributes of the incumbents as a means of describing the job? Despite this disagreement, many researchers seem capable of agreeing on at least a partial definition of what constitutes a job analysis.

McCormick (1976) included in his definition of job analysis the collection of information on the following components: (a) "job-oriented" behaviors including tasks and work procedures; (b) "worker-oriented" behaviors such as providing supervision and making decisions; (c) working with machinery and equipment; (d) performance evaluation criteria such as error rates and productivity; (e) work environment factors; (f) and personnel requirements such as personality characteristics, physical abilities, and skills. Harvey (1991), while largely agreeing with McCormick's definition, takes exception to the personnel requirements listed in (f).

According to Harvey (1991), job analysis should "be applied only to procedures that *collect* information describing verifiable job behaviors and activities; it should *not* be used to denote the wide assortment of procedures that make inferences about people or otherwise *apply* job analysis data…"

Observables vs. Inferred Characteristics

At the heart of this issue is the question of whether or not job analysis should include components that are not directly observable or at the least, indirectly observable through what McCormick et al. (1972) referred to as "strong inference." According to Harvey (1991), the three main goals of job analysis are the description of observables, the description of work behaviors that are independent of the worker, and collecting data that is replicable and verifiable. Personality characteristics and human abilities do not fit these requirements because they require large degrees of inference, and, in essence, they change the orientation of job analysis. When one makes inferences of required human attributes, the level of analysis ceases to be the job and becomes the person instead.

Despite this, many popular job analysis instruments seem to disregard this prescription and consider the collection of such inferential information an indispensable part of the process. Indeed, the *Standards for Educational and Psychological Testing* (American Psychological Association, 1999) and the Society for Industrial and Organizational Psychology's (SIOP) *Principles for the Validation and Use of Personnel Selection Procedures* (1987) both include the designation of abilities or worker specifications as legitimate inclusions in job analyses.

Job Analysis vs. Job Specification. Immersed in the debate over whether or not human traits and abilities are appropriate for inclusion in job analyses is the need to distinguish between job analysis and job specifications. Recent literature has zeroed in on this very theme. Harvey and Wilson (2000) take issue with Sanchez and Levine (2000) due to what they consider to be the latter's use of the term job analysis to refer to "any activity that involves making ratings regarding work activities or the prerequisite worker-traits needed for successful job performance" (pp. 835). It is stressed that job analysis is the process of collecting information about what is done on the job, whereas job specification is the process of inferring the human traits or abilities required for successful job performance. One – job analysis – should necessarily precede the other – job specification. Although many researchers and analysts seem to blur or completely overlook this distinction (see Sanchez and Levine, 2000; Fleishman and Reilly, 1992; and Fleishman and Mumford, 1991), the *Uniform Guidelines on Employee Selection Procedures* (1978) are very specific on the issue. The *Uniform Guidelines* define job analysis as the examination of work behaviors and then note that while work behaviors may contain both "observable (physical) and unobservable (mental) components"; however, "…Knowledges, skills, and abilities are not behaviors…"

Consistent with the ongoing debate as to what comprises a job analysis and what does not, the numerous commercially available job analysis instruments that exist are also divided as to the type of data they attempt to collect. Whereas some collect observable data, other instruments collect only information of worker attributes, and still other job analysis instruments side step the issue by including a universe of information – both observable and inferred.

It is ironic that a procedure so readily used and referred to suffers from so much controversy and has so many "unknowns" associated with it. Despite the facts that job analyses are an essential feature of almost every activity engaged in by industrial-organizational psychologists, that the EEOC's *Uniform Guidelines* place such a strong emphasis on the need for accurate job analysis as the necessary means of establishing job relatedness, and that the courts have shown great deference to the *Guidelines* in rulings on job relatedness (e.g., Griggs v. Duke Power Co., 1971; and Albemarle Paper Co. v. Moody, 1975) many questions central to job analysis have gone unexamined.

Job Analysis Philosophies and Instruments

Task-Oriented Philosophies. Historically, job analysis has been centered on obtaining information that is very technologically and behaviorally specific. These are often referred to as "task-oriented" or "job-oriented" (McCormick, 1976) philosophies.

With these instruments data is collected that centers on tasks that result in some specific, observable outcome. Task oriented job analysis may be applied to one job at a time, or with the use of task inventories, several different jobs may be analyzed utilizing the same questionnaire. Obviously, these instruments take a large amount of time and human resources to complete; however, a large amount of information that can be readily applied to personnel functions such as selection or development of training programs is the outcome.

The FJA. The Functional Job Analysis (FJA; Fine, 1971) is an example of a task-based job analysis instrument. It is typically applied to one job at a time and results in a very detailed list of tasks tailored to the job. The FJA attempts to identify exactly what a worker does on the job as well as the results of the worker's behaviors. Included in the information are not only what gets done (the actual task) but also to whom or what, why the task is performed, how the task is performed (including what types of equipment or tools are used and what types of instruction or guidance are provided), and the relative amount of orientation toward data, people, or things involved in the task.

Task inventories have been heavily used by government agencies, the military, and by very large corporations for the purpose of collecting job relevant information in a manner that is inexpensive and less time consuming than are traditional task analyses (see Gael, 1983 & The U.S. Department of Labor, 1973 for examples). The premise behind task inventories is that respondents judge whether or not the position of interest is reflected in each item of the questionnaire. In a task inventory, only some subset of tasks listed will apply to the job in question (as the inventory is assumed to include the entire universe of tasks for a given study); tasks that are deemed relevant are usually "checked-off" of the list. In some instances, and in order to help meet the *Guidelines'* requirement for identifying behaviors that are essential to performance of the job, the applicable tasks are then rated in terms of frequency of performance, importance, judged difficulty, criticality or some other category that quantifies the task at hand.

Although rich in detailed information, task or job-oriented job analysis techniques are often considered lacking in many respects. Because the emphasis is necessarily on the job, the relativistic ratings that are made with most of these instruments are not suitable for the purpose of comparing task-dissimilar positions that may have other underlying similarities or requirements.

Worker-Oriented Philosophies. Worker-oriented philosophies of job analysis have been proposed as one means of addressing these limitations and facilitating cross-job comparisons. Worker-oriented philosophies of job analysis seek to describe jobs on a common set of descriptors that are more behaviorally and technologically abstract than are tasks (McCormick, 1976). According to McCormick et al. (1972), the basic premise of worker-oriented job analysis is that there exists a relatively small set of general work behaviors that apply to all jobs, and that one can describe all jobs in terms of how much of each of these general work behaviors are involved.

As pointed out by Harvey (1991), the distinction between what is a worker-oriented and what is a task-oriented statement is not always clearly marked. Theoretically, for each worker-oriented statement in an inventory, one should be able to identify several discreet tasks that demonstrate the performance of the worker-oriented item. However some worker-oriented items are considerably more specific than others, and might be equally well considered task-oriented items. Likewise, no standard exists against which one may decide if the statements are too broad or holistic to be rated accurately. Whereas the unit of measurement is more generalized work behaviors or activities, there is little agreement on how general is too general. The critical distinction between the two philosophies lies, therefore, not in the actual wording of the items, but in the purpose of the analysis.

The PAQ. Numerous worker-oriented instruments currently exist for one to choose among. One of the most widely used and researched instruments is the Position Analysis Questionnaire (PAQ; McCormick et al., 1972). The PAQ consists of 194 items or job elements that fall into the following categories: (a) information input (where and how the worker gets the information he/she uses for the job); (b) mental processes (reasoning, planning, decision making, etc. involved in the job); (c) work output (the physical activities the worker performs and the tools or devises used); (d) relationships with other persons; and (e) job context (physical and social environment in which the work is performed). The individual items provide for checking an item if it applies, or for rating it on an appropriate rating scale such as importance, time, or difficulty.

Although the average item reliability of the PAQ has been found to be .80, it has been noted that the questionnaire has some faults. The first issue centers on the domain of jobs appropriately covered; the PAQ appears most suited for use with blue-collar or manufacturing jobs, as opposed to professional, managerial, or some technical jobs (Cornelius et al., 1984; DeNisi et al., 1987). Secondly, since the PAQ does not describe specific work activities, behavioral similarities in the jobs may mask genuine task differences between them – for example, a police officer's profile is quite similar to a housewife's (according to Arvey & Begalla, 1975) because of the troubleshooting, emergency-handling orientation required by both positions. Additionally, readability is an issue with the PAQ; it requires a college-graduate reading level in order to comprehend the items (Ash & Edgell, 1975). Hence, the PAQ should not be given to job incumbents or supervisors unless their jobs require educational levels substantially higher than 10 to 12 years. Finally, the PAQ has been criticized on the basis of the potential for it to be utilized in such a matter that cross-job comparisons are untenable due to the potential for raters making within-job-relative ratings (Harvey, 1991). Instructions for making ratings may encourage respondents to rate importance only with respect to the job's other applicable tasks, and not in relationship to other jobs (assumably those that we truly wish to be able to make comparisons to).

The JEI. In an effort to make a worker-oriented instrument more widely applicable, the Job Element Inventory (JEI; Cornelius & Hakel, 1978) was developed. By editing the PAQ, a 153-item questionnaire utilizing only one rating scale (Relative Time Spent) was developed. This new questionnaire had a much lower reading level than the PAQ – 10th grade, and factor analytic studies indicated that the JEI shared a very similar factor structure with the PAQ (Harvey et al., 1988). Although an obvious improvement in regard to reading level, the other criticisms of the PAQ still stand for the JEI.

The F-JAS. Focusing largely on the theory that the best way to analyze a job is in terms of the personal characteristics and traits required of the incumbents, the Fleishman Job Analysis Survey was developed (F-JAS; Fleishman, 1975; 1992; Fleishman & Reilly, 1992). Its objective is to describe jobs in terms of the abilities required to perform the jobs. The abilities-requirements taxonomy is intended to reflect the fewest independent ability categories that describe performance in the widest variety of tasks.

Areas covered by the F-JAS taxonomy include: (a) 21 cognitive abilities (e.g., problem sensitivity, flexibility of closure, spatial organization, and selective attention); (b) 10 psychomotor abilities (e.g., arm-hand steadiness, multi-limb coordination, and response orientation); (c) 9 physical abilities (e.g., dynamic flexibility gross body equilibrium, and explosive strength); and (d) 12 sensory abilities (e.g., glare sensitivity, auditory attention, and sound localization) (Fleishman et al., 1999). Because the items to be rated are quite broad and holistic in nature, rating scales that define each ability, distinguish it from related abilities, and provide examples of tasks that require different levels of the ability are used to facilitate understanding and reliable ratings among the respondents as well as provide a framework allowing comparisons across different jobs.

Evidence of the F-JAS's comprehensiveness has been reported in several studies. Hogan et al. (1979) found that 80% of the tasks performed by warehouse workers could be assigned to one or more of the ability categories. Similar findings are reported in regard to Army officers (Mumford et al., 1985), FBI special agents (Cooper et al., 1983), and New York City police officers (Landy, 1988). Reports of reliabilities for the instrument vary across the positions and types of respondents used - from .66 (Romashko et al., 1976) to in excess of .90 when using 15 or more judges for several different jobs (Hogan et al., 1978). Despite the large amount of research that has been conducted on the F-JAS, it has its critics among those industrial-organizational psychologists that believe that inferred human characteristics are not the correct unit of measurement for job analysis because they are not directly observable and require broad inferences.

The CMQ. A relatively new job analysis instrument that seeks to address the dilemmas faced by some of the other worker-oriented instruments, while still enabling the comparison of task dissimilar jobs with a broad worker-oriented philosophy, is the Common Metric Questionnaire (CMQ; Harvey, 1990). CMQ ratings are made on absolute scales that describe how much of a work activity is present, such that meaningful cross-job comparisons are possible.

The CMQ was written using an eighth grade reading level and includes items that are equally appropriate to executive, professional, and managerial positions as to non-managerial, manufacturing, or blue-collar jobs (Harvey, 1993). Consistent with Harvey's (1991) prescription that job analysis be concerned with the measurement of "observables", the CMQ was designed such that it includes items that are more behaviorally abstract than tasks, but less behaviorally abstract than some of the other, more holistic instruments that include unobservable job and person characteristics.

The CMQ collects over 2000 item-level ratings that can be combined to describe jobs in terms of 80 work dimensions (a holistic component), or when greater specificity is required, the decomposed, item-level data can be used (Harvey, 1993). Despite the numerous ratings composing each of the 80 work dimensions, median coefficient alpha for the scales is .86 (Harvey, 1993).

*The O*Net*. The most recent, and perhaps ambitious, addition to the spectrum of worker-oriented job analysis instruments is the U.S. Department of Labor's Occupational Network, or O*Net. This instrument is currently being utilized by the U.S. government for the purpose of collecting data on thousands of positions as part of an ongoing research endeavor aimed at replacing the Dictionary of Occupational Titles (DOT; U.S. Department of Labor, 1991).

Because of the highly specific, task-level data collected and presented in the DOT, it was determined that the DOT was unsuitable for current needs. Namely it could not be used for rapid assessment of the skills, knowledges, and other characteristics required by a job family or for showing how skill levels in one job may relate to those of other jobs (Dunnette, 1999). It was decided that the new occupational information system, the O*Net, would need to possess occupational information allowing for cross-job comparisons, information that includes the tasks and behaviors involved in the work and the conditions under which the work is performed, the requirements imposed on incumbents, and attributes arising from experience (such as skills and expertise) as well as the more basic individual attributes (including abilities, interests, and personality characteristics). In order to cover this huge domain, the O*Net is divided into nine different scales; each designed to measure one aspect of interest.

The first scale, Basic and Cross Functional Skills, refers to the capabilities that are required to do the job of interest. In particular, skills, unlike abilities, are learned and amenable to change and are usually considered observable through what McCormick (1972) termed "strong inference." This taxonomy is composed of 46 items making up seven lower-order skills including: (a) content skills (e.g., reading comprehension and speaking); (b) process skills (e.g., learning strategies and critical thinking); (c) complex problem-solving skills (e.g., information organization and implementation planning); (d) social skills (e.g., persuasion and service orientation); (e) technical skills (e.g., operations analysis and troubleshooting); (f) systems skills (e.g., identification of downstream consequences and systems evaluation); and (g) resource management skills (e.g., management of financial resources and management of personnel resources) (Mumford et al., 1999). Each item is rated on the absolute level of the skill, the relative performance of the skill, and any job specialty requirements can be noted.

The initial validation study for the O*Net included 648 incumbents and supervisors who provided ratings for the Skills scale (Mumford et al., 1999). Interrater agreement for the level scales ranged from .75 to .92, with a median of .84. For the importance scale, ***r***s ranged from .71 to .93 with a median of .83. In all cases (and for all O*Net scales described herein) interrater agreements were computed including "does not apply" or "not relevant" ratings; despite the fact that prior studies have suggested that these ratings should be excluded because they may inflate correlations (Harvey & Hayes, 1986).

The second scale of the O*Net is the Knowledges scale. As it applies to this scale, knowledge is defined as a collection of discrete but related facts and information about a particular domain (Costanza et al., 1999). The Knowledges taxonomy is comprised of 33 different items broken into 10 different types of knowledge: (a) business and management (e.g., economics and accounting and personnel and human resources); (b) manufacturing and production (e.g., production and processing and food production); (c) engineering and technology (e.g., building and construction and mechanical); (d) mathematics and science (e.g., physics and psychology); (e) health services (e.g., medicine and dentistry and therapy and counseling); (f) education and training; (g) arts and humanities (e.g., foreign language and philosophy and theology); (h) law and public safety (e.g., public safety and security and law, government, and jurisprudence); (i) communications (e.g., telecommunications and communications and media); and (j) transportation. Ratings are made on an absolute level scale, relative importance scale, and job specialty requirements can be designated for each item.

Interrater agreements for the Knowledges scale from the initial studies on the O*Net (N=645 incumbents and supervisors for this scale) ranged from .37 to .95 for the level scale and from .55 to .93 for the importance scale (Costanza et al., 1999).

The third scale of the O*Net is the Occupational Preparation scale (Anderson, 1999). This scale includes information on the education, training, experience, and licensure/certification requirements associated with a job. The ratings that are obtained include: (a) general education level (rated on a 12-point degree-level scale); (b) instructional program (check the discipline that applies); (c) subject area education level (for each of 15 descriptors rate on a 5-point scale); (d) licenses required (yes/no rating for two descriptors plus listing of specific licenses); (e) requirement to obtain a license (six descriptors rated yes/no); (f) who requires the license (three descriptors rated yes/no); and (g) related work experience (four descriptors rated on a 11-point scale).

The initial validation study of the O*Net (N=598 for this scale) reports interrater agreements that range from .08 (for several of the instructional program ratings) to .97 (for general educational level) for the Occupational Preparation Scale (Anderson, 1999).

The fourth scale of the O*Net is the Generalized Work Activities (GWA) scale (Jeanneret et al., 1999). This scale is most consistent with the previous conceptualizations of job analysis in that it is intended to reflect the work and behaviors that are actually performed on a job; it is for all practical purposes a very broad worker-oriented task list. The GWA includes 41 items that are broken down into nine different subscales. Many of the items included in this scale are derived from other job analysis instruments, including the Position Analysis Questionnaire (McCormick et al., 1972), the Occupational Analysis Inventory (Cunningham, 1988), the General Work Inventory (Cunningham et al., 1990), and the Job Element Inventory (Cornelius and Hakel, 1978). These include: (a) looking for and receiving job-related information (e.g., getting information and monitoring processes); (b) identifying and evaluating job-related information (e.g., identifying objects, actions, and events and estimating the characteristics of materials, products, events, or information); (c) information data processing (e.g., judging the qualities of objects, services, or persons and processing information); (d) reasoning/decision making (e.g., thinking creatively and organizing, planning, and prioritizing work); (e) performing physical and manual work activities (e.g., controlling machines and processes and performing general physical activities); (f) performing complex/technical activities (e.g., interacting with computers and implementing ideas, programs, systems, or products); (h) communicating/interacting (e.g., establishing and maintaining interpersonal relationships and performing for or working directly with the public); (i) coordinating/developing/managing/advising others (e.g., developing and building teams and providing consultation and advise to others); and (j) administering (e.g., staffing organizational units and monitoring and controlling resources) (Jeanneret et al., 1999).

Like the other O*Net scales reviewed thus far, the GWA is necessarily holistic in orientation; however, there is some degree of variability with regard to how specific the items are. Ratings for the GWA scale originally included the absolute level, relative importance, and absolute frequency of the work activities; current forms have discontinued the frequency scale due to time requirements for administration. Interrater agreement for the GWA level scale ranged from .51 to .92 with a median of .82. For the importance scale, ***r***s ranged from .59 to .92 with a median of .68 (Jeanneret et al., 1999).

The fifth scale of the O*Net is the Work Context scale (Strong et al., 1999). This scale is designed to measure the environment in which work is accomplished, including applicable social, physical, and organizational contexts that may impact the worker or serve as the impetus for tasks and activities. Many job analysis instruments routinely collect information of this type.

Thirty-eight items make up the Work Context scale that is factored into two first-order factors and 10 second-order factors. The primary factors are: (a) interpersonal relationships (e.g., communication and conflictual contact with others) and (b) physical work conditions (e.g., job demands and pace and scheduling). The scales utilized for rating these items varied across items and included absolute level, absolute frequency, relative importance, likelihood of occurrence, and seriousness of injury. Across all items the range of interrater agreement was .20 to .97 with a median of .83 (Strong, et al., 1999).

Organizational Context is the sixth scale of the O*Net (Arad et al., 1999). Information about organizations is expected to be of interest because it is believed that the nature of jobs will vary as a function of the characteristics of the organizations in which they occur. In particular this information is expected to give insight into the relationships between certain business practices and effectiveness (i.e. “high performance organizations”).

Unlike other O*Net scales, the Organizational Context scale is aimed primarily at obtaining information on organizations, not on jobs. Over 70 items were administered to either a single organizational HR representative, an incumbent, or both (Arad et al, 1999). Items were designed to measure two types of structural characteristics – organizational structure (e.g., employee empowerment and job characteristics) and human resources systems and practices (e.g., group socialization and recruitment operations) and four types of social processes: (a) organizational culture and values; (b) goals (e.g., individual goal specificity and extent of organizational goal specificity); (c) roles (e.g., role conflict and role negotiability); and (d) leadership (e.g. consideration and visionary), as well as miscellaneous high-performance constructs (e.g. use of independent contractors and change in organizational structure).

Because of the data collection methodology, analyses of the Organizational Context scale reliabilities were conducted for the incumbents and HR representatives separately. For incumbents, the interrater agreement coefficient across occupations ranged from .16 to .83 and across organizations from .04 to .62. Because only one HR representative from each organization was interviewed, internal consistency reliability coefficients were computed. Coefficients ranged from .67 to .88 (Arad et al., 1999).

Factor analyses of this data provided support for the phenomena of high-performance organizations. Elements found to be included in this domain were: decentralization, information sharing, use of teams, use of data in decision-making, risk-taking values, and multiple-skills training (Arad et al., 1999).

The seventh scale of the O*Net, Abilities, is arguably a controversial one, as it seeks to measure human characteristics that many have suggested are inappropriate in a job analysis context. For purposes of the O*Net, and consistent with many psychological definitions (Carroll, 1993; & Fleishman, 1975), abilities are seen here as being traits or relatively enduring attributes of an individual's capabilities for performing a particular range of different tasks.

The abilities included in the O*Net scale are the same abilities measured by the Fleishman Job Analysis Survey (F-JAS) (Fleishman, 1975) and include 52 items (see earlier discussion of the F-JAS for a breakdown and examples). All 52 items are rated on a seven-point absolute level scale as well as a job-relative importance scale. Results from the O*Net validation study yielded interrater agreement indices for the level scale ranging from .45 to .90 and for the importance scale from .33 to .92.

Occupational Interests and Values represent the eighth scale of the O*Net (Sager, 1999). This information is collected for the purpose of aiding in one of the major goals of the O*Net, namely, person—job matching. A large body of literature is centered on increasing employee satisfaction and performance through what is often called "person—organization fit" (Hakel, 1986; Dawis, 1991; Borgen et al, 1968).

In order to describe the interest domain of occupations, it was decided to utilize Holland's six-factor taxonomy (Gottfredson & Holland, 1996; Holland, 1976) for the O*Net (Sager, 1999). Holland "codes" can be used to describe an occupation in terms of one, two, or three Holland types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. The three codes are associated with each job in the order of importance. For the purposes of the O*Net the codes were rated as interests of incumbents. As these codes were not collected via incumbent/supervisor questionnaires for the O*Net, reliability analyses were not available.

In order to measure occupational values, the Minnesota Job Description Questionnaire (MJDQ; Borgen et al., 1968; Dawis, 1991; Dawis & Lofquist, 1984) was selected because it was believed that the twenty-one reinforcers (e.g., ability utilization, advancement, authority, independence, recognition, security, variety, and working conditions) measured by this instrument provided reasonable coverage of the work-values domain. Each of the twenty-one reinforcers is rated on a 5-point Amount of Agreement scale. Reliabilities for the reinforcers range from .10 to .79.

The final scale of the O*Net, Work Styles, centers on personality requirements associated with jobs (Borman et al., 1999). Like the Abilities scale previously discussed, this scale has the potential for controversy as it sets out to measure constructs that are not directly observable nor identifiable via "strong inference" (McCormick, 1972). The Work Styles scale is designed to measure seven first-level constructs along with 17 second-level constructs. First level constructs with corresponding second-level constructs are as follows: (a) achievement orientation (achievement/effort, persistence, and initiative); (b) social influence (energy, leadership orientation); (c) interpersonal orientation (cooperation, concern for others, social orientation); (d) adjustment (self-control, stress tolerance, adaptability/flexibility); (e) conscientiousness (dependability, attention to detail, integrity); (f) independence (independence); and (g) practical intelligence (innovation, analytical thinking) (Borman et al., 1999). Each of the preceding is rated on an absolute level scale, as well as a relative importance scale. Reliabilities for level ratings range from .16 to .84, with a median of .66. Reliabilities for importance ratings range from .26 to .84 with a median of .64.

The O*Net exhibits great potential if it is able to live up to its very ambitious goals for describing jobs. As with the other instruments reviewed here, its success will be largely dependent on its ability to show not only reliable and accurate ratings, but also strong external validity.

Relativity of Rating Scales.

An issue of relevance to both philosophies of job analysis is to what degree do the rating scales used permit cross-job comparison. That instruments utilize job-relative rating scales is a criticism of many existing job analysis techniques, not just task-oriented scales. Although worker-oriented instruments strive to describe jobs such that broad behavioral similarities may be evidenced, rather than obscured by task dissimilarity (Harvey, 1991), some worker-oriented instruments still fall short of this goal as a result of relying upon with-in-job relative rating scales.

According to Harvey (1991), "if the rating scales, instructions to raters, or subsequent score transformations cause a job analysis item's numerical rating to be expressed *relative to other items performed on that job*, the item ratings will be deemed to be only with-in job relative" (pp. 82).

Unfortunately, when relativistic ratings such as this occur, it may not only be impossible to make cross-job comparisons, it may be impossible to make cross-position comparisons for individuals holding the same job title. Relative ratings are only able to convey that for a given job some task is performed more or less frequently than other tasks for that same job (or is more or less important, or is more or less critical, etc.). Ultimately this information, while perhaps interesting, lacks much with regard to utility.

Current Issues in Job Analysis

The proliferation of job analysis instruments over the past two decades is not only the result of advances in technology and differing philosophies; changes have also come about as a means of addressing concerns expressed by organizations regarding the time and labor intensive nature of most job analysis instruments. The desire to make instruments amenable to multiple positions, collecting numerous types of data, and yet less intrusive and time/labor intensive has challenged industrial-organizational psychologists for several years.

Addressing the Issues

Who Should Rate? One method aimed at decreasing the intrusive nature and the time demands of the job analysis process has been to utilize individuals who may not have extensive familiarity with the job or the job analysis instrument. Smith and Hakel (1979) is one of the most frequently cited articles for supporting the claim that naïve job analysts provided with only job titles can rate as accurately as those given extensive job descriptions. Using the Position Analysis Questionnaire, ratings were obtained from incumbents, supervisors, job analysts, and college undergraduates for numerous government jobs. The undergraduates were provided with either a job title only, or with job titles plus written job descriptions. On the basis of very high convergent validity correlations (.89 to .98) obtained, Smith and Hakel determined that the source and amount of information used by analysts made "little practical difference" (1979, p. 691).

Jones et al. (1982) is another frequently cited reference supporting the concept of using job analysis raters with only nominal job information. College students made PAQ ratings after having received only short narrative job descriptions. Results were reported as being reliable, valid, and highly consistent with findings obtained during the initial PAQ validation efforts (Jones et. al, 1982).

Cornelius et al. (1984) also obtained high interrater agreement indices (average r = .52) when using students to provide job ratings for nine positions with the PAQ. Students were provided with job titles only. However, convergent validities with expert ratings were lower than anticipated, and when validities were recomputed using only tasks that the experts had identified as "Non-Does-Not-Apply," the validities dropped further (from .58 to .41). Cornelius et al. (1984) ultimately concluded that the ratings provided by job naïve job analysts were not equivalent to those provided by expert raters utilizing the PAQ.

Harvey and Hayes (1986) took a much closer look at the "Does Not Apply" issue as it might relate to inflating job analysis reliabilities and validities on the PAQ. Of concern was whether or not high reliability on obviously irrelevant items (which can be considered useful information in and of itself) may combine with low reliability on the "Non-Does-Not-Apply" items resulting in the incorrect assumption of acceptable overall interrater reliability. Harvey and Hayes' results indicated that reliabilities in the .50 range could be obtained when raters rule out only 15-20% of the items as "Does Not Apply" – even when the remaining items are responded to in an entirely random manner. These results caused Harvey and Hayes to conclude, "the interrater correlations reported for relatively job-naïve PAQ raters by Jones et al. (1982) and Smith and Hakel (1979) are dangerously low" (1986; pp. 356). Furthermore, it is noted that interpretation of this data should serve as "evidence…against the use of untrained or relatively job-naïve raters…in job analysis applications" (pp. 356).

In another test of whether or not raters with severely limited amounts of job-related information could provide accurate PAQ ratings, Friedman & Harvey (1986) provided job-naïve raters with varying amounts of job descriptive information and compared obtained reliabilities and validities with those of job content experts. Analyses conducted both including and excluding Does Not Apply items consistently yielded unacceptable reliability coefficients (means ranged from .41 to .57), as well as unacceptably low convergent validities (means from .41 to .56). It was concluded that even when provided with short job descriptions, job naïve raters are incapable of providing the same quality of PAQ ratings as more expert analysts.

Although it seems fair to conclude that job-naïve raters are not the best source of ratings for the PAQ, the question remained as to how job-naïve raters might function on a non-PAQ type of job analysis instrument. Harvey and Lozada-Larsen (1988) examined the accuracy of such raters when utilizing a 521-item task inventory. College undergraduates provided ratings for eleven insurance-related positions that were then compared to ratings obtained from actual incumbents. Results indicated that participants using either a job title and description or a description only were considerably more accurate than those utilizing a job title only. Once again it was concluded that employing job-naïve raters as a means of reducing the costs of job analysis is a dubious strategy (Harvey & Lozada-Larsen, 1988).

Level of Specificity. Another, and perhaps more promising, method of reducing the time demands and intrusiveness of job analysis is to move toward more holistic worker-oriented instruments. Not only are such instruments now being implemented (e.g. the O*Net), but it is a strategy that has been recommended as one means of adapting the very popular PAQ so it can be more widely used (Friedman & Harvey, 1986). Despite the promising nature of this endeavor, very few researchers have sought to examine the reliability and validity of ratings obtained from the more holistically oriented instruments.

Cornelius and Lyness (1980) compared the accuracy and reliability of holistic and decomposed job ratings for 10 different occupations. In the decomposed condition, participants were given a task inventory and then asked to rate each task based upon 13 different scales. In the holistic condition, participants were given a copy of the task list to look over and review, but were then required to make holistic ratings of the job on the 13 scales (not on a task-by-task level). Interrater agreements were highest, on average, for the decomposed judgment strategy (mean ***r***s of .54 vs. .40). Hierarchical regressions using the degree of agreement between incumbents and job analysts and supervisors yielded results that were not statistically significant; however results were in the same direction as the interrater consistency measures. Cornelius and Lyness (1980) thus conclude that the two techniques provided “equally good job analysis ratings” (pp. 161).

Sackett et al. (1981) sought to determine the degree to which holistic ratings of job similarity would coincide with a significantly more time-consuming task based analysis of job overlap for the purposes of job classification. Eight foreman jobs in a chemical processing plant were compared utilizing either a task inventory or a paired-comparison technique. Hierarchical cluster analyses and multidimensional scaling analyses both indicated that the two techniques produced comparable results. The Pearson correlations between the two were .84. These results were consistent with other studies that have utilized holistic judgments strictly for the purpose of grouping jobs (e.g., Pearlman et al, 1980); however the authors do note that "the advantage of this technique over others is the time and cost savings in the job analysis phase of a research project…This comparison may be labeled unfair in that the task analysis yields a larger body of data which may be used for other purposes" (pp. 801).

Consistent with the above, Cornelius, Schmidt, et al. (1984) found that individuals could utilize job title information only to correctly classify jobs into one of four a priori job families. Analyses indicated that classifications obtained from the holistic judgments were comparable to those obtained using discriminant analysis of highly decomposed rating profiles.

Heeding his own advice, Butler and Harvey (1988) compared holistic and decomposed ratings of the Position Analysis Questionnaire. Three groups of raters (professional job analysts, graduate students in industrial-organizational psychology, and PAQ-unfamiliar undergraduates) made direct holistic ratings of the PAQ for four familiar jobs. These holistic ratings were then compared to decomposed ratings that had been made by the professional job analysts. For the holistic ratings, interrater reliabilities were as follows: professional job analysts: $r = .49$; graduate students: $r = .43$; and for undergraduates; $r = .26$. Utilizing the Cronbach Accuracy procedure, Butler and Harvey (1988) report exceedingly low levels of holistic rating accuracy for all groups of interest. They conclude that making holistic job analysis judgments likely overloads the cognitive capacity of the participants.

The most recent effort to examine the accuracy of holistic job analysis ratings was conducted by Sanchez and Levine (1994). Their study included three jobs and had participants either rate an entire task inventory on each of 12 dimensions (decomposed condition), or to simply rate the 12 dimensions as they applied to their occupation (holistic condition). Dependent variables included interrater agreement, root mean square deviation from experts' consensus ratings, and correlation with experts' consensus ratings. The use of decomposed ratings received support for only one of the three jobs considered when interrater agreement was used as the dependent variable (although two of the three jobs exhibited reliabilities in the predicted direction, only one was statistically significant). However, rating decomposition significantly increased accuracy for two of the three jobs. When correlation was the dependent variable, the results showed a trend in the predicted direction. It was concluded that, especially for lower level positions, the decomposed strategy of making job analysis ratings results in more accurate and reliable data.

Because of the fact that the few studies that have examined the efficacy of holistic measures have obtained mixed results, this area remains a major gap in our knowledge with regard to job analysis. Unlike most task inventories, holistic instruments require raters to consider the job as a whole, rather than rate individual components. This requires complex cognitive processes of combining information about all aspects of the work to make inferences about several presumed underlying dimensions of the work. Disagreement over the reliability and validity of such job analysis instruments stems largely from our lack of understanding the information processing demands of such rating tasks.

Cognitive Decision Making

Almost all texts on cognitive information processing begin with the "rule of seven." This is based on early research by Miller (1956) showing how severely limited the information processing capacities of humans are. Essentially, for most types of information that we need to receive, process, or remember, the average human's capacity is seven pieces of information, plus or minus two. Unfortunately, all commercially available job analysis instruments require more information than this, but instruments that involve more decomposed tasks may be more successful because the decomposed nature of the ratings lessens the emphasis on memory.

Consistent with Miller's prescription (1956), Bethell-Fox and Shepard (1988) have recently shown that increasing the complexity of information causes increasing difficulty in cognitive processing. Specifically they found that the time for encoding, mental rotation, and comparison making were increased as the complexity of the stimulus increased. The implications of this for holistic job analyses may be found in the trade off with regard to time. Although the complex nature of the ratings may cause participants to take longer to make each rating than they would on a more decomposed instrument, the significantly fewer items to be rated may still leave holistic instruments ahead with regard to time consumption.

In order to effectively deal with the increased demands of processing complex information, subjects may seek ways of simplifying the situation cognitively (Einhorn, 1971). Although cognitively "breaking-down" the amount of information required was found to be the most frequently used strategy of individuals facing complex mental tasks, it appeared that decision makers then did not utilize an additive model to reconstitute the information for rating purposes. Ultimately it was concluded that accuracy was negatively impacted for judgments requiring large amounts of information.

Research from the Clinical Arena. Much of the research concerning holistic vs. decomposed judgments can be found in the clinical psychology arena. Studies aimed at analyzing the accuracy of clinical prediction have shown that clinicians are most likely to contribute effectively through mechanically combining pieces of discrete information as opposed to making holistic judgments (Sawyer, 1966).

Goldberg (1967; 1968) likewise reported that clinical judgments tend to be "rather unreliable" and "rather low in validity on an absolute basis." He found through a series of studies in physical medicine (diagnosing malignant vs. benign ulcers), psychiatry (deciding on advisability of granting temporary liberty to psychiatric patients), clinical psychology (diagnosing neurosis vs. psychosis), and securities analysis (predicting future stock prices) that having judges make inferences about decomposed cases and then combine them via simple linear modeling increased judgment quality.

Einhorn (1972) also looked at the accuracy of holistic judgments made within the clinical domain. She had three pathologists predict survival on the basis of information contained in biopsies taken from patients with cancer. Einhorn concluded that prediction of this type is vastly improved by having judges mechanically combine numerous components of information taken from the biopsies as opposed to making a global judgment on survival.

Taken as a whole, these studies point toward the effectiveness of having raters make numerous judgments regarding their jobs and then mathematically compiling them to yield an accurate job description.

Research on Decision Making. The other major source of information on the information processing demands of complex information is the decision-making literature.

Streufert el al. (1965) examined the changes that occur in information search and information utilization as a result of information complexity. Their findings indicated that both information search and utilization appear to decrease as a result of the cognitive load associated with highly complex information.

Stimuli composed of many separable elements (i.e. holistic rating tasks) may result in more extreme ratings than the sum of their constituent parts (Manis et al, 1966). Three experiments requiring judges to evaluate social stimuli that were composed of multiple parts were conducted. Although the results were consistent with a model that equates a compound stimuli as equal to a weighted average of its constituents, a large effect was found for the number of constituents in each stimuli; all other factors being equal, the number of elements was directly related to the extremity of the judgment made.

Research suggests that information processing systems may respond in a curvilinear fashion to three components of input load: complexity of information, unpleasantness, and pleasantness (Driver & Streufert, 1969). Consistent with this, several studies have shown a negative impact on information processing effectiveness and performance as complexity of information is increased.

Jacoby et al. (1974) examined the results of supplying consumers (undergraduate students) with increasing amounts of information on laundry detergents. Data indicated that increasing the amount of information available on the laundry detergent packages beyond a certain point resulted in dysfunctional consequences in terms of the consumer’s ability to select the brand of detergent to best fit his/her needs.

In an extension of the earlier work, Jacoby, Speller, and Berning (1974) operationalized increasing information as the number of brands and amount of information per brand for rice and T.V. dinners. Results from housewives that participated confirmed the prior finding that human beings are only able to assimilate a finite amount of information during any given time period, and once that limit is surpassed, behavior becomes confused and dysfunctional.

Armstrong et al. (1975) also looked at the decrements in performance that accompanied highly complex materials. Students were divided into two groups, one assigned to solve a decomposed version of a problem and the other to solve a holistic/direct version of the problem. The results provided support for the hypothesis that people can make better judgments when they use the principle of decomposition, and that decomposition is especially valuable in situations or problems where participants have little knowledge of the subject matter.

A study examining the effectiveness of securities analysts examined the success of decisions based on the amount of information utilized (Jacoby et al., 1985). Findings indicated that the better performing participants generally considered slightly greater amounts of information and different types of information than did the poorer performing analysts. These results were interpreted in the context of availability of information resources to draw upon when making decisions and the issue of whether or not one has control over the amount and types of information consulted when making decisions.

MacGregor et al. (1988) had participants make quantitative judgments utilizing a mathematical algorithm under five different conditions of aid (ranging from highly structured to no aid). Results showed increasing amounts of accuracy and consistency as the nature of the task became more decomposed and highly structured. Improved accuracy was also seen when making quantitative judgments of probability distributions utilizing decomposed as opposed to holistic ratings (Hora et al., 1993).

Decision Making in Academia. The final area where comparisons of holistic and decomposed judgment strategies have been made is in the academic arena – specifically in the ratings of applicants to graduate programs or in the ratings of professors.

Dawes (1971) reported that in general, a linear combination of multiple applicant credentials is superior in predicting candidate success in a graduate program than are committee's holistic determinations of candidate adequacy.

Keeley and Doherty (1972) found that when four PhDs were asked to make admissions judgments on hypothetical applicants, multiple regression analyses out-performed the professors in regard to predicting who would be selected for admission.

When students made ratings of hypothetical college professors, interrater reliabilities across occasions, interrater convergence across rating methods, and interrater agreements were all found to be superior in the decomposed judgment vs. holistic judgment strategy (Lyness and Cornelius, 1982). However, correlational analyses indicated that a simple holistic strategy was as effective as the decomposed approach. These mixed findings caused the authors to conclude that while a strategy of combining decomposed ratings was most effective, the holistic strategy is acceptable.

All of the above studies seem to indicate that if our desire is to increase accuracy of job analysis ratings, a superior strategy is to utilize ratings of a more decomposed nature. Having individuals deal with overly complex cognitive decisions (those requiring numerous, coincident judgments) may result in completely ineffectual behavior, whereas providing structure or decomposing the judgment into smaller successive judgments may increase performance considerably. However, the few studies that have actually been conducted in the job analysis domain have yielded mixed results. Furthermore, despite the fact that it makes use of very holistic ratings, no studies of this nature have been conducted on one of the largest job analysis projects in history – the O*Net.

Current Study

Although no one study can solve the many ills of job analysis, this study aims to examine two current issues in the practice and the research – are differing sources of job analysis ratings comparable to one another, and to what degree are holistic job ratings accurate when compared with more decomposed ratings of work behaviors. Of course, in order to do this, it is of theoretical consequence to consider whether or not it is even appropriate to speak of the "accuracy" of job analysis ratings.

Conceptualizations of Job Analysis Accuracy

Sanchez and Levine (2000), in response to the large amount of literature detailing human fallibility in judgment making, have taken the position that the accuracy of job analysis can not be obtained by comparisons with some standard. Consequential validity is argued to be the true area of concern, or put another way "the accuracy of the job analysis data is defined by their effects, and it cannot be determined in the absence of knowledge of such effects" (pp. 813).

In a similar vein, Morgeson and Campion (2000) adopted the stance that considering accuracy as the deviation from some standard is untenable in light of the fact that "there are rarely unambiguous standards against which to judge" (pp. 822).

Disputing these assumptions, Harvey and Wilson (2000) stated that the view that there is no objective reality with regard to job analysis ratings is only likely if one (a) takes a highly relativistic/metaphysical view of reality (Is there any reality at all?), or (b) insists on confusing the accuracy of job analysis with the accuracy of inferences associated with job specification. Consequential validity is concerned with documenting the correctness of inferences involved in the application or use of job analysis data, not the data themselves.

From a practical standpoint it could be argued that the potential for making quality inferences from inaccurate data is very slim; just ask the last person who purchased a "cream-puff" that turned into a "lemon" based on the used car salesperson's "valid" assessment of the car's quality. From the standpoint of psychological science, the view that *all* accuracy is a socially constructed phenomenon must be rejected; how does one ever make headway if one cannot reject some theories as being wrong – despite the fact that they possessed utility. Research exploring potential confounds in obtaining accurate job analysis ratings is imperative – not only for the sake of improving our instruments or rating methods, but also for legal defensibility of the inferences we must make when utilizing job analysis information for other purposes.

Potential Sources of Job Analysis Inaccuracy.

Drawing from the domains of Social and Cognitive psychology, Morgeson and Campion (1997) delineated a framework of sixteen potential sources of inaccuracy in job analysis.

Social Sources of Inaccuracy. The social sources of inaccuracy that are postulated to be of interest are broken down into social influence processes and self-presentation processes. Included under the domain of social influence are pressures toward conformity with a group, group polarization of opinions, and social loafing or motivational loss. One self-presentation process to be wary of included impression management, or engaging in behaviors to create and maintain desired perceptions (Schlenker, 1980). Social desirability responding is a second area of potential concern. It is the process where people attempt to gain approval by engaging in culturally appropriate and acceptable behaviors (Marlowe & Crowne, 1961). The final self-presentation process listed concerns the impact of demand characteristics, or aspects of the research situation that may cause participants to try and be “good” by guessing and validating the researcher’s hypotheses (Orne, 1962).

The social sources listed are suggested to have differential impacts on job analysis based upon who provides job analysis ratings, the setting in which analysis ratings are collected, the type of job analysis instrument utilized, and the purpose for which the analysis is conducted (Morgeson & Campion, 1997).

Cognitive Sources of Inaccuracy in Job Analysis. Both limitations in cognitive information processes and biases in cognitive information processes are suggested to have the potential for a profound impact on the accuracy of job analysis instruments.

Information overload is potentially applicable to all types of raters in numerous situations. It has been repeatedly shown that when confronted with exceedingly complex information, subjects will engage in simplification strategies that also result in decrements in accuracy (see Einhorn, 1971; reviewed above). Simplification strategies may include the use of heuristics (mental shortcuts), or categorizations (stereotypes).

Inaccuracy in job analysis judgments may also result from either intentional or unintentional biases in information processing. Carelessness is one such cognitive bias that may evidence itself either as a result of the people making the ratings or the nature of the rating task. Having to process irrelevant, extraneous information in a job analysis setting is also prone to cause bias. Prien and Saleh (1963) hypothesized that factors such as tenure and job difficulty would bias analyst ratings; and Arvey et al. (1982) found that the degree of enthusiasm exhibited by incumbents biased analyst ratings.

Inadequate information is also postulated to be a source of bias in job analysis ratings. This is consistent with studies of naïve job raters reviewed earlier (Cornelius et al., 1984; Friedman & Harvey, 1986; Harvey & Lozada-Larsen, 1988).

Halo, leniency, and severity are all highly researched biases, both in the performance appraisal and social psychological domains. Inaccuracy from halo occurs when ratings or judgments are assigned on the basis of global impressions or highly salient features. Halo has been shown to impact sentencing in court procedures and to result in "spill-over" or high or low performance ratings (social reference needed; Borman, 1975, 1977). Leniency, or the tendency to give consistently high ratings, has been shown to have a large impact in performance appraisals. The desire to not be overly critical or downgrade a position may result in incorrectly high ratings (Benedict & Levine, 1988; Bernadin & Beatty, 1984), and self-ratings of performance (similar to incumbent ratings of jobs) appear to be especially prone to inflated ratings (Harris & Schaubroeck, 1988; Mabe & West, 1982). Severity is not anticipated to impact job analysis ratings (Morgeson & Campion, 1997).

Method effects are another class of biases researchers are advised to be aware of as potential sources of inaccuracy (Morgeson & Campion, 1997). In general, method effects are expected as a result of individuals attempting to make ratings that are consistent with previous ratings of a stimulus, or because responses to prior items orient the individual's attention to certain responses – priming (Pfeffer, 1977).

Job analysis appears to be fertile ground for inaccuracy resulting from both cognitive processing limitations and biases. The rating task is necessarily complex, requiring large amounts of recall and integration of information; making numerous fine distinctions further predisposes ratings to inaccuracy by laying the ground for methods effects in the data to emerge. In addition, it is human nature to attempt to present the self in a valuable light and to associate overly positive traits with jobs that we value and enjoy.

The foregoing social and cognitive processes have been repeatedly demonstrated to exist and influence judgments in numerous situations; it seems very likely that they also exist and have the potential to exert influence in the job analysis arena. The current study will examine several hypotheses relating to the potential for inaccuracy in job analysis ratings. Although the results will not provide direct insight into the causes of inaccuracy, findings will help to generalize the proposed inaccuracy framework (Morgeson & Campion, 1997) to the domain of job analysis.

The framework presented by Morgeson and Campion (1997) provides little definitive insight into what type of job analysis rating schemes should be most accurate -- holistic or decomposed, or what type of rater should be most accurate (incumbents, professional analysts, or job-naïve); instead it puts forth a framework that may be used for making competing hypotheses.

Hypotheses Related to Holistic vs. Decomposed Job Analysis Measures

Several aspects of social and cognitive psychology lead to the expectation that decomposed job analysis ratings are likely to be more accurate than are holistic ratings.

Self-presentation Processes. Social desirability responding is much more likely to occur with instruments that are highly abstract in nature as opposed to those that are more decomposed (Smith and Hakel, 1979). This is appears to result from the fact that holistic job analysis instruments allow respondents to rate work activities in what they believe is the “right” way without having to substantiate ratings at a more discreet level.

Limitations of Information Processing Systems. In job analysis, making holistic judgments is more complex than making judgments of more specific or decomposed tasks because larger amounts of recall and integration of information is required. When information overload occurs, it seems likely that people will engage in simplifying strategies as a means of handling the information.

One such strategy for simplifying information is the use of heuristics. Heuristics are essentially shortcuts that allow us to judge objects (or tasks) in terms of how representative of a category they are, and how available the information is to us (Nisbett & Ross, 1980; Tversky & Kaahnerman, 1974). Although highly functional in day-to-day exchanges, heuristics are often highly inaccurate judgment strategies.

Categorizations are another means by which humans simplify their world. In particular, when one has little information to use for a frame of reference, one often relies on stereotypes or assumed notions of "what belongs" in order to make decisions (Wyer & Srull, 1981). Holistic job analysis ratings by their very nature provide respondents with very little information.

Biases in Information Processing Systems. Halo biases may result when one makes judgments on broad, global impressions (Thorndike, 1920; Borman, 1991). Questionnaires that are highly abstract or have non-specific descriptors are likely to result in overlapping and ill-defined ratings (Cooper, 1981). In addition, the potential for halo to bias ratings is more pronounced if categorization has already been activated. Categorization results in an over-all judgment of the entire job that then spills across all ratings.

In the opposite vein, past research and social/cognitive theories can also be found that support the hypothesis that more holistic ratings of worker behaviors should be more accurate than are decomposed ratings.

Social Influence Processes. The failure to participate to one's best ability is often attributed to lack of motivation. Not only can lack of motivation be associated with issues like social loafing, but it may also arise due to characteristics of the task. Job analysis questionnaires that are overly long, uninteresting, or anonymous may result in lack of motivation. And as in the case of any type of performance, lack of motivation typically results in decreased outcomes. Participants who find themselves dealing with lengthy, decomposed job analysis instruments may very well lose motivation to complete it accurately.

Limitations of Information-Processing Systems. Information overload not only occurs when making highly complex judgments, it can also occur when the amount of information is very large. Many decomposed job analysis instruments require ratings of thousands of items on numerous dimensions. In order to deal with this large amount of information, raters may very well respond in a highly redundant manner as a means of reducing the cognitive workload (Friedman, 1990; Sanchez & Frazer, 1992). This redundancy has the potential to lead to inaccuracy in ratings.

Biases in Information Processing Systems. Carelessness, or unintentionally responding inaccurately, inappropriately, or missing items is another type of processing bias. Green and Veres (1990) found that incumbents making task ratings were more inclined toward carelessness as length of task list increased. Similarly, Wilson et al. (1990) found careless responding was less problematic on shorter instruments, and that careless responding negatively impacted reliabilities. In particular, when highly decomposed task lists contain numerous “Does Not Apply” ratings (as they necessarily will), respondents may become bored and inattentive.

While each side of the holistic/decomposed argument may be valid, consideration of the prior research in the job analysis domain leads to the expectation that a decomposed job analysis instrument will be more accurate than is a more holistic instrument. By comparing both types of instruments to an external job analysis of the same position, and measuring convergence (operationalized as the correlation) with this criterion one can assess the accuracy of the given job analysis. Furthermore, as interrater agreements have been stated to be the “standard means of assessing the reliability” (Butler & Harvey, 1988) of job analysis instruments, computations of these will shed light on the degree to which raters report engaging in the same work behaviors on the two types of instruments. The following hypotheses are proposed:

***Hypothesis 1*: A decomposed job analysis instrument will yield more accurate job analysis ratings than will a more holistic instrument**.

***Hypothesis 2*: Interrater agreement will be higher on a decomposed job analysis instrument than on a more holistic job analysis instrument.**

Hypotheses Related to Source of Job Analysis Ratings

Several studies previously reviewed examined the issue of accuracy with regard to who makes job analysis ratings. Although the findings and methodologies utilized seem to be predisposed toward Expert Job Analysts, the Morgeson and Campion (1997) framework provides the means to make several different hypotheses here, as well.

Social Influence Processes. The presence of others is not necessary for them to influence our behavior – this is one of the central tenets of social psychology. If individuals typically work as part of a team, or within a group, the group's norms and ideals will typically influence members on tasks that are not completed within a group framework. Because many organizations have taken-up a team-based approach in recent years, it is safe to assume that many incumbents do have a strong social identity with their working peers. This may result in pressures toward conformity in ratings, as opposed to accuracy. It is expected that when compared with incumbents, job analyst experts will be less susceptible to these social pressures because of their more objective outsider position (Morgeson & Campion, 1997).

Self-presentation Processes. Because impression management behaviors are aimed at having others view us in a particular light (Schlenker, 1980), completing job analysis instruments provides an opportunity for incumbents to provide information that may cause others to view them as highly competent, industrious, and perhaps even deserving of a raise.

Social desirability responding is a second self-presentation process that incumbents are likely to feel pressure toward. Often organizational culture dictates how a job is seen and incumbents may be motivated to distort ratings as one method of portraying their position in a more socially desirable light. Likewise, certain tasks may be rated as especially important, while other, lower level tasks are played down (Arnold & Feldman, 1981).

Unlike incumbents, job analysis experts have nothing to gain by engaging in self-presentation strategies that will distort job ratings. Their tendency to engage in behaviors that might inflate the ratings of a job should be very low.

Limitations of Information-Processing Systems. Job analysis rating tasks are very likely to be a new experience for incumbents. Regardless of the specificity of the instrument used, the rating task will probably require considerations of large amounts of information, and perhaps even conceptualizations of the job in a new manner. The cognitive demands of such thought processes will likely result in information over load and the accompanying decrements in performance. Job analysis experts, on the other hand, are typically very experienced in the job analysis process and are accustomed to dealing with large amounts of job relevant data for numerous types of positions. Furthermore, familiarity with the process lessens the potential for feeling overwhelmed by the rating process and the opportunity to focus on obtaining accurate data.

Biases in Information Processing Systems. Incumbents rating jobs is analogous to individuals making self-ratings in any domain. When making self-ratings, the tendency toward leniency is great (Harris & Schaubroeck, 1988; Mabe & West, 1982). Therefore, one should expect greater degrees of inaccuracy among incumbent ratings of job analysis than among expert ratings.

In general, one could argue that job analysis experts will be less prone to social influence, self-presentation tendencies, biases such as leniency, and better equipped to handle the cognitive demands of job analysis.

Limitations in Cognitive Processing Systems. When motivation is questionable, the reliance on mental short cuts such as heuristics is likely to increase. Students typically have low investment and little incentive to engage in systematic processing when participating in campus research due to the lack of accountability.

Social influence processes. Students can typically not be held accountable for their responding, and it is questionable whether or not they find the nature of the research project meaningful. The large groups that are assembled for the purpose of collecting data may also stimulate social-loafing as they are not task interdependent nor do they contribute largely to the student's social identity (Smith & Mackie, 1999).

As before, consideration of the aforementioned literature does not point toward a clear path for obtaining accurate job analysis ratings. However, when considered in relation to what we have found in prior job analysis studies, the following is expected and will once again be tested based on convergence with an external job analysis:

> ***Hypothesis 3*: Expert job analysts will provide the most accurate job analysis ratings and job-naïve raters will provide the least accurate job analysis ratings.**

Limitations in Information Processing Systems. Utilization of heuristics or other mental short-cuts when dealing with complex information is often a function of prior knowledge or experience with some stimuli. For job analysis experts, it would not be unusual to have rated numerous similar positions over time, and as such have developed expectations for the job at hand. This prior experience is a likely precursor for activation of not only representativeness heuristics, but also anchoring and adjustments. The impact of these last two has been examined in past job analysis literature (Smith et al., 1989); findings indicated that after having been exposed to a job title, this job title served to anchor expectations for the job at hand and often resulted in making insufficient adjustments when provided with stereotype inconsistent information. When job analysts automatically process information on positions, they are also likely to rely on categorizations that may obscure true differences between the job being rated and existing stereotypes resulting in inaccurate information.

Self-Presentation Processes. Demand characteristics of the situation are likely to impact the behavior of both job-naïve raters (undergraduates) and incumbents in a similar direction. By inadvertently conveying to participants the direction in which ratings would be preferred (i.e. if a job analysts states the purpose of job analysis as being to develop a selection procedure, incumbents might infer that ratings of higher complexity are desired as a means of insuring qualified hires), researchers might cause incumbent raters to respond in a more similar manner than might otherwise occur. Job analysts should be relatively free of motivation to perform in this manner as they are usually considered the authority figure in job analysis situations.

Information Processing Biases. The amount of information held by raters has been shown to directly impact the reliability of job analysis ratings (Hahn & Dipboye, 1988). Incumbents are naturally expected to have considerably more job relevant information when making ratings than will either job analysis experts or job-naïve raters.

Limitations in information processing, self-presentation and information processing biases all appear to have the potential to influence to what degree raters respond in a similar manner. Reliance on stereotypes is most likely for those with limited job information, and greater job information will likely include knowledge of true differences in work behaviors. Inasmuch as this is true, interrater agreements should reflect this with higher levels of agreement for those relying upon stereotypes and lower levels of agreement for those with more direct job knowledge. As such, the following is anticipated:

Hypothesis 4: **Expert job analysts will show the greatest degree of interrater agreement and job incumbents will show the least amount of interrater agreement.**

Hypotheses Related to the Issue of Complexity

Furthermore, the role of complexity impacting accuracy in the decision-making literature is applicable to the job analysis research at hand. Numerous studies have indicated that as the complexity of a decision increases the reliability (measured as both agreement and temporal reliability) and validity (criterion and convergent) of the decision decreases (Armstrong et al., 1975; Fischer, 1976, & Morera & Budescu, 1998). In the job analysis domain it is typically assumed that holistic measures, by their very nature, require individuals to make numerous simultaneous attribute judgments, and are thus very complex in nature. However, in order to ensure this level of complexity it would be necessary to actually present raters with numerous attributes that needed to be considered at the same time. This way, rather than simply assuming raters are considering numerous relevant elements when making a holistic judgment, one can be more certain of this as the facets are actually presented to the rater for consideration. In order to assess this in a direct manner it was necessary to develop a holistic scale that actually included many work behaviors associated with each holistic item and then compare this with the standards of interest. As such, interrater agreements and convergence with an external criterion will be assessed for what should be a truly cognitively complex task.

***Hypothesis 5*: When the degree of complexity is increased in the job analysis rating process, the accuracy of ratings will decrease.**

***Hypothesis 6*: When the degree of complexity is increased in the job analysis rating process, the level of interrater agreement will decrease.**

Hypotheses Regarding the Combination of Rater and Instrument Type

As current trends in practice focus on both the type of instrument used and the type of raters utilized, consideration of the combination of the two is justified. Drawing from the hypotheses outlined above with regard to rater type and instrument type, the convergence of each type of rater's ratings on each type of instrument with an external job analysis will yield information on the overall accuracy of each.

***Hypothesis 7:* Expert job analysts utilizing a more decomposed job analysis instrument will provide the most accurate job analysis ratings.**

***Hypothesis 8:* Job-naïve raters utilizing a more holistic job analysis instrument will provide the least accurate job analysis ratings.**

Although previous studies have examined similar hypotheses in regard to other job analysis instruments, there are several novel elements to the current study. First, the instruments utilized in the current study are actually two different instruments that both purport to measure the same constructs or work behaviors accurately, however, one utilizes very broad holistic items and the other uses much more specific, decomposed items. Prior studies have typically used the same instrument broken into holistic and decomposed components and have only compared the ratings obtained to each other, not an external criterion. Second, this is the first study (to the author's knowledge) that sets out to compare the Department of Labor's O*Net to an external criterion as a means of judging its accuracy. As the O*Net is designed to replace the current DOT and be utilized for selection, training, and other personnel functions at a national level, it is imperative for both legal and ethical reasons that the data collected by it be accurate. Finally, this study actually manipulates level of complexity as a means of determining if prior assertions that increased complexity cause increased inaccuracy in job analysis ratings are warranted.

Method

Participants

The participants for this study were comprised of three different groups – the position incumbents, expert job analysts group, and the job-analysis-naive group.

The position incumbent group included 35 individuals currently holding the position of Secretary within an organization. The organizations represented included three universities, a real estate office, and two car dealerships. All secretaries reported having never completed a job analysis before.

Consistent with prior studies of this type, graduate students in industrial-organizational psychology and professors of industrial-organizational psychology from two universities constituted the expert job analysts group (N = 34). Only those graduate students having had course work in job analysis and/or having had experience actually carrying out job analytic procedures were utilized.

The job analysis-naïve group consisted of 120 undergraduate students from two universities who participated for extra-credit. Only those students who had not had course work in industrial and organizational psychology nor had held the position of secretary for a period of time exceeding two months were utilized to help ensure their "naïve" status.

Materials

All participants were provided with a narrative job description of the target position "Secretary" (see Appendix A) as well as general instructions for completing the two job analysis instruments. This position was chosen because it is expected to be familiar to all study participants, and was shown via pilot testing to be familiar to undergraduates as well as graduate students in industrial-organizational psychology in a prior study (Butler & Harvey, 1989).

The Generalized Work Activities Questionnaire (GWA), Form B, of the O*Net (U.S. Department of Labor & National O*Net Consortium, 1999) was the first of two job analysis instruments used in this study. The GWA is one of nine scales that comprise the O*Net and was utilized here because the content it measures is consistent with the emphasis of almost all previous job analysis techniques and instruments – to collect information on the work that is done on the job. The GWA functioned as the holistic measure of job analysis for the study; it consists of 41 holistic work activities, and has participants make ratings on both the relative importance and the absolute level of the various work activities. Participants rated the position of interest utilizing a web-based version of the O*Net scale. Appendix B contains a paper and pencil version of the scale. The web-based version of this instrument is identical to the paper and pencil version.

In order to consider the issue of complexity in holistic ratings, a second form of the O*Net scale was developed for web-based responding. This form was identical to the original with one difference – the inclusion of all representative work behaviors derived from the CMQ. Following the definition of each GWA, respondents then saw a "For example…." listing the associated behaviors that corresponded to the item of interest. Respondents did not rate these exemplars, they served only to increase the number of attributes that had to be simultaneously considered for each item; thus increasing the cognitive complexity of the holistic rating. Half of the undergraduate participants completed each of the two versions of the O*Net.

The Common-Metric Questionnaire (CMQ) (Personnel Systems and Technologies Corporation, 1993) served as the decomposed job analysis instrument used for data collection in this endeavor. The CMQ is a worker-oriented instrument that allows for the description of jobs in terms of 80 work dimensions that are computed from mathematical aggregation of item level data (Harvey, 1993). In addition to overlapping the nine O*Net dimensions, the CMQ provides information on numerous work behaviors not measured by the O*Net GWA scale. The CMQ was administered by providing participants with a computer floppy disk containing the computer-adaptive job analysis program. Appendix C contains a copy of the paper and pencil version of the CMQ.

Procedure

GWA ratings. All groups rated the target position of Secretary utilizing the Generalized Work Activities scale (with half of undergraduates each responding to the less/more complex version). In addition to receiving written instructions, all participants were given verbal instructions on the use of the instrument. Prior to rating the target position, participants were given the opportunity to make a “practice” rating of a different (non-secretary) position for one of the generalized work activities.

CMQ ratings. After having completed the GWA ratings, all groups rated the target position utilizing the CMQ. It was decided to administer the CMQ after the GWA for two reasons. First, the CMQ was given after the GWA to prevent the more decomposed nature of the CMQ from causing priming effects in the ratings on the GWA (Pfeffer, 1977). Second, this would be consistent with the practice of how a job analysis utilizing a holistic instrument would be carried out. One of the primary reasons for utilizing holistic instruments is to save time and therefore money. In order to save time individuals completing such an instrument would not complete a longer, more behaviorally specific instrument first; at best they would be given the opportunity to review a job description and proceed from there.

Experimental Design

Because the CMQ is designed to measure a considerably larger number of work behaviors, it was necessary to match the dimensions measured by the two instruments before making most comparisons. Even at the dimension level, the CMQ is more decomposed than is the O*Net and dimensions did not necessarily match up one to one.

For example, whereas the O*Net has one dimension for "Communicating and Interacting with Others" the CMQ has 24 dimensions that cover multiple forms of internal contacts, external contacts, and attending and chairing meetings. Appendix D shows how the two scales were matched (utilizing all O*Net dimensions and the appropriate subset of CMQ dimensions) and the resulting six dimensions that are used for this study.

Although some have argued that the best measure of job analysis accuracy is the degree to which it is utilizable, convergent validity is the approach taken here to measure accuracy. Accuracy was operationalized as the correlation between obtained job analysis ratings and an external criterion set of job analysis data. This criterion data was gleaned from a considerably larger, more heterogeneous data set of job analyses of various secretarial and administrative type positions. The first step in trimming the original 400+ cases was to remove those positions that were obviously highly divergent from the position rated here in the work performed. Specifically, a large number of ratings were for positions that were either considerably less or more complex than the position of interest (e.g. they were of positions that were more "clerk" like or "executive administrator" in orientation). The second step included removal of those ratings that were obviously done in a careless or inaccurate manner (those raters that marked numerous items as either "Apply" or "Does-Not-Apply" incorrectly). The final step in creating the criterion was purely statistical in nature. From the remaining 115 incumbents' ratings individuals were removed one at a time in such a manner as to maximize the interrater agreement among the criterion data set. The final criterion included ratings from 50 secretarial incumbents utilizing the CMQ with an overall interrater agreement = .7698.

Interrater agreement was assessed in order to determine the similarity between the profiles obtained using each type of instrument and provided by the different types of raters. According to Butler and Harvey (1989) "average interrater ***r***s are a standard means of assessing the reliability of job analysis ratings and indicate the degree of similarity between the profiles provided by raters in each group." They are computed by finding the ***r***s between all possible pairs of raters, converting these to ***z*** equivalents, averaging across raters in the group, and converting back to ***r*** units.

For both the CMQ and the O*Net interrater agreement indices were computed for the entire scale as well as for the six dimension scores described previously by rater type.

Results

Accuracy Hypotheses

In order to test the hypotheses related to accuracy, the convergence between the ratings provided by participants in this study and an external job analysis of the position secretary was examined. Convergence was operationalized as the correlation between the job analysis of interest and the criterion. Because neither job analysis instrument included any dimension that all raters rated as "Does Not Apply," ratings for all dimensions are included in the analyses.

Hypothesis 1 stated that a decomposed job analysis instrument would yield more accurate job analysis ratings than would a more holistic instrument. The correlation between the CMQ and the criterion data set is considerably larger than the correlation between the O*Net and the criterion data set, thus supporting this hypothesis (r = .573 vs. r = -.241) (see Tables 1 & 2).

Hypothesis 3 stated that expert job analysts would provide the most accurate job analysis ratings and that job-naïve raters would provide the least accurate ratings. Examination of correlations alone would indicate this was true for the CMQ data (r = .675 vs. r = .451), as well as for the O*Net data (r = .256 vs. r = -.154 & -.258). As a follow-up to this, tests for significant differences in correlations by rater type were conducted. Appendix E shows the results of T tests for these differences. For the CMQ it was found that no two raters' correlations were significantly different from each other, and as such it is not appropriate to state that any one rater was more accurate than another for the decomposed instrument. It was expected that on the O*Net scale that had been altered so as to increase the level of complexity (by providing respondents with numerous attributes to consider simultaneously), rating accuracy would be decreased. O*Net job-naïve raters utilizing this more complex version of the instrument made ratings least converging with the criterion data set (r = -.258) (see Tables 1 & 2); and their correlations were found to be significantly different from those of both incumbents and analysts [t = 1.365(df = 93), p = .05 & t = 2.112 (df = 92), p = .05]. These data only provide partial support for Hypothesis 5 because although raters utilizing the complex measure had correlations that appeared to converge with the criterion to the least degree, it was not shown that their ratings were significantly different from other job naïve raters [t = .4583 (df = 118)]. Consideration of both correlations and differences among correlations ultimately suggests that for holistic instruments the choice of rater can make a significant difference in accuracy. Nevertheless, it cannot be determined unequivocally who makes the most accurate ratings when utilizing a decomposed job analysis instrument – significant mean differences did not exist between raters.

Hypotheses 7 and 8 stated that job analysts utilizing a decomposed job analysis instrument would provide the most accurate job analysis ratings and that job-naïve raters utilizing a holistic job analysis instrument would provide the least accurate job analysis ratings, respectively. Both hypotheses were supported (r = .675 and r = -.154 and -.258).

Table 1 shows correlations between the CMQ and criterion data set; Table 2 presents correlations between the O*Net and criterion data set. Both are computed based on the total scores for each instrument. Appendix E contains convergence correlations at the dimension level for the instruments.

Table 1. Correlations between Criterion Data and CMQ Data.

	CMQ Incumbents N = 35	CMQ Analysts N = 34	CMQ Job-Naïve Raters N = 120	CMQ - All Raters Included N = 186
Criterion Data Set N = 50	.582*	.675*	.451*	.573*
Standard Errors	.164	.186	.142	.129

* p=.05

Table 2. Correlations between Criterion Data and O*Net Data.

	O*Net Incumbents N = 35	O*Net Analysts N = 34	O*Net Job-Naïve Raters N = 60/60[a]	O*Net - All Raters Included N = 186
Criterion Data Set N = 50	.122	.256	-.154 / -.258[a]	-.241
Standard Errors	.165	.172	.137 / .120	.114

[a] Job-naïve raters utilizing the more complex GWA rating scale.

Interrater Agreement Hypotheses

Interrater agreements were computed by first taking the correlation between all possible pairs of raters on all items making up each dimension. The average of these correlations was then computed and serves as the interrater agreement coefficient reported here. As stated before, no dimension (or item composing the dimension) of either the CMQ or the O*Net received “Does Not Apply” ratings from all respondents. As such, interrater agreement indices reported below were computed using all dimensions. Consistent with Peterson et. al (1999) interrater agreements for the O*Net scale were computed on “Importance” and “Level” items separately for each dimension.

Hypothesis 2 stated that interrater agreement indices would be higher on a decomposed job analysis instrument than on a holistic instrument. This hypothesis received only partial support (see Table 3). At the scale level interrater agreements for each type of rater, as well as the average interrater agreement across all raters, were higher for the CMQ than for the O*Net. However, at the dimension level there was a great deal of variability in regard to where the highest levels of interrater agreement were seen based upon the respondent type. Table 3 contains interrater agreements for all raters on both instruments.

Expert job analysts were predicted to show the greatest degree of interrater agreement and job incumbents were predicted to show the least amount of interrater agreement by Hypothesis 4. This was supported at the scale level for the CMQ, and on five of the six scale dimensions. The hypothesis was partially supported for the O*Net scale; although analysts did show the highest interrater agreement on the O*Net, job-naïve raters showed the least amount of agreement. This trend was also displayed at the dimension level for the O*Net (see Table 3).

Hypothesis 6 indicated that for raters making holistic judgments, increased complexity would result in lower interrater agreements. This was marginally supported at the scale level, and supported on three of the six dimensions (see Table 3).

Although no dimension was rated as “Does Not Apply” by one hundred percent of raters, it was still felt to be important to consider the possibility that simply having a large number of “Does Not Apply” ratings could artificially inflate the interrater agreements obtained. In order to examine this the number of items within each dimension rated as “Does Not Apply” was obtained and correlated with the interrater agreement indices. Significant correlations were found for both Dimension 5, “Decision Making with Regard to Company Resources” ($r = .3643$, $p<.01$) and Dimension 3, “Working with Technical Equipment and Performing Technical Activities” ($r = .4538$, $p<.01$).

Table 3. Interrater Agreements for Criterion Data Set, the CMQ and the O*Net Dimensions and Full Scales.

Rater Type	Incumbents N = 35	Analysts N = 34	Job-Naïve Raters N = 60/60[a]	Averaged Across Rater Type
Dimension 1				
Criterion Data Set				.5549
CMQ	.1927	.3398	.2032	.2452
O*NET Importance Ratings	.1425	.2704	.0275 / .1180[a]	.1396
O*Net Level Ratings	.0274	.2646	.0519 / .0490[a]	.0982
Dimension 2				
Criterion Data Set				.5651
CMQ	.1978	.3784	.2499	.2754
O*Net Importance Ratings	.4120	.6402	.2890 / .3145[a]	.4139
O*Net Level Ratings	.3390	.5656	.2890 / .3262[a]	.3780
Dimension 3				
Criterion Data Set				.9855
CMQ	.4218	.8185	.6737	.6380
O*Net Importance Ratings	.8629	.7665	.7580 / .6577[a]	.7613
O*Net Level Ratings	.7690	.7232	.6038 / .5889[a]	.6712
Dimension 4				
Criterion Data Set				.9820
CMQ	.0070	.6532	.4097	.3566
O*Net Importance Ratings	.1172	.2816	.2459 / .2272[a]	.2180
O*Net Level Ratings	.3057	.5149	.3837 / .3224[a]	.3817
Dimension 5				
Criterion Data Set				.1344
CMQ	.2296	.4025	.1453	.2591
O*Net Importance Ratings	.4382	.5949	.2893 / .2505[a]	.3932
O*Net Level Ratings	.3609	.5507	.3058 / .2406[a]	.3645
Dimension 6				
Criterion Data Set				.9385
CMQ	.2756	.5988	.3836	.4193
O*Net Importance Ratings	.3204	.4268	.1792 / .2250[a]	.2879
O*Net Level Ratings	.2821	.3947	.1098 / .2432[a]	.2575
Full Scale				
Criterion Data Set				.7698
CMQ	.4388	.5909	.5093	.5130
O*Net Importance Ratings	.4239	.5876	.3277 / .3219[a]	.4153
O*Net Level Ratings	.3782	.5352	.3003 / .2969[a]	.3777

[a] indicates job-naïve raters utilizing the more complex holistic rating instrument.

Although the results cited above address the hypotheses made for this study, it was felt that further analyses could further illuminate issues of interest. These analyses are discussed below.

Differences in Dimension Ratings Due to Rater Type. In order to further examine the degree to which different types of raters provided comparable job analysis ratings, differences in mean ratings for each of the six dimensions on both the O*Net and the CMQ were tested for. For the O*Net, Dimensions 1 and 6 showed significant differences by rater type (see Table 4). Post hoc analyses indicated that statistically significant differences between incumbents and analysts ($t = 2.170$, $p = .05$) and incumbents and job naive ($T = 1.099$, $p = .05$), respectively, were at issue. For the CMQ, significant mean differences were found for Dimension 6 (see Table 5); post hoc analyses indicated that the significant differences existed between incumbents' and job naive ratings ($t = -7.536$, $p = .05$). Figure 1 shows the average ratings for each dimension by rater type for the O*Net; Figure 2 shows this for the CMQ. Appendix F contains dimension means and standard deviations for each scale.

Table 4. Mean Differences by Rater Type for O*Net Dimensions

	Sum of Squares	df	Mean Square	F	Sig.
O*Net 1	82.675	3	27.558	5.916	.001
O*Net 2	14.899	3	4.966	1.725	.163
O*Net 3	3.217	3	1.072	.574	.633
O*Net 4	22.821	3	7.607	2.616	.052
O*Net 5	22.196	3	7.399	2.283	.081
O*Net 6	27.260	3	9.087	3.465	.017

Table 5. Mean Differences by Rater Type for CMQ Dimensions

	Sum of Squares	df	Mean Square	F	Sig.
CMQ 1	2.504	2	1.252	.376	.687
CMQ 2	18.621	2	9.311	.223	.800
CMQ 3	134.365	2	67.183	.560	.572
CMQ 4	201.603	2	100.802	1.181	.309
CMQ 5	2.550	2	1.275	.015	.985
CMQ 6	1487.023	2	743.511	3.650	.028

Convergence of Dimensions across Scales. Examination of the dimensions measured by the O*Net and the CMQ indicated that the two overlapped and Appendix D indicates how they were matched for the purposes of this study. In order to ascertain to what degree the dimensions actually converged, correlations were computed between the dimensions of interest. A review of Table 6 indicates that the within each scale most dimensions were significantly correlated; however, across the two scales the dimensions are typically *not* significantly correlated, and occasionally show negative correlations. The only exception to this was on dimension one, "training, managing and coordinating others" ($r = .160$, $p = .01$).

As it might be possible to argue that the low convergence between the CMQ and O*Net dimensions presented here is due to the method in which they were matched up, and that they do in fact actually measure the same work behaviors it appeared prudent to examine to what degree one could successfully predict any of the CMQ dimension scores utilizing all of the O*Net dimensions, and vice versa. Multiple Rs for each dimension on both the CMQ and the O*Net were computed utilizing all six of the dimensions from the opposing job analysis instrument. A review of Table 7 indicates that in no instance was the variance accounted for significant.

Table 6. Correlations among O*Net Dimensions Scores, CMQ Dimensions Scores, and Criterion Dimension Scores.

	OD2	OD3	OD4	OD5	OD6	CD1	CD2	CD3	CD4	CD5	CD6	CRD1	CRD2	CRD3	CRD4	CRD5	CRD6
OD1	.683*	.360*	-.013	.611*	.520*	.160*	.063	-.015	.048	.031	.152*	-.119	-.191	-.144	-.995	.134	-.252
OD2	1.000	.345*	.081	.605*	.476*	.044	.025	-.018	-.028	-.003	.104	-0.15	-.316*	-.183	-.135	.004	-.213
OD3		1.000	.028	.464*	.332*	-.002	.043	-.099	-.065	-.002	-.020	-.209	-.835*	-.172	-.262*	-.108	-.179
OD4			1.000	.053	-.008	.077	-.049	.018	.029	-.001	.026	.019	.117	-.26	-.682*	.127	-.991*
OD5				1.000	.616*	.106	.063	-.080	.044	-.001	.101	-.154	-.958*	-.143	-.144	.074	-.168
OD6					1.000	.018	.044	-.046	.011	-.017	.111	-.149	-.124	-.109	-.183	-.304*	-.173
CD1						1.000	.716*	.335*	.415*	.549*	.442*	.591*	.556*	.651*	.636*	.268	.555*
CD2							1.000	.349*	.373*	.594*	.387*	.824*	.651*	.539*	.587*	.361*	.865*
CD3								1.000	.431*	.172*	.370*	.405*	.696*	.431*	.736*	.665*	.769*
CD4									1.000	.193*	.639*	.654*	.789*	.504*	.317*	.246	.474*
CD5										1.000	.433*	.307	.263	.706*	.569*	.423*	.687*
CD6											1.000	.679	.535*	.828*	.737*	.647*	.486*
CRD1												1.000	.590*	.138	.529*	.584*	.538*
CRD2													1.000	.061	.328*	.491*	.532*
CRD3														1.000	.363*	-.035	.288
CRD4															1.000	.283*	.555*
CRD5																1.000	.372*
CRD6																	1.000

* significant at the .01 level

*Key: OD → O*Net Dimension #; CD → CMQ Dimension #; CRD → Criterion Dimension #.*

Table 7. Multiple Rs for each dimension of the O*Net and CMQ when predicted by all six dimensions from the opposing instrument.

Dependent Variable	R	df	F	Sig.
O*Net D1	.142	6	.550	.770
O*Net D2	.169	6	.786	.582
O*Net D3	.114	6	.352	.908
O*Net D4	.257	6	1.881	.087
O*Net D5	.157	6	.674	.670
O*Net D6	.180	6	.892	.502
CMQ D1	.131	6	.468	.831
CMQ D2	.109	6	.319	.926
CMQ D3	.198	6	1.084	.374
CMQ D4	.100	6	.269	.951
CMQ D5	.205	6	1.175	.322
CMQ D6	.080	6	.171	.984

Figure 1. Mean Dimension Ratings for the O*Net by Rater Type

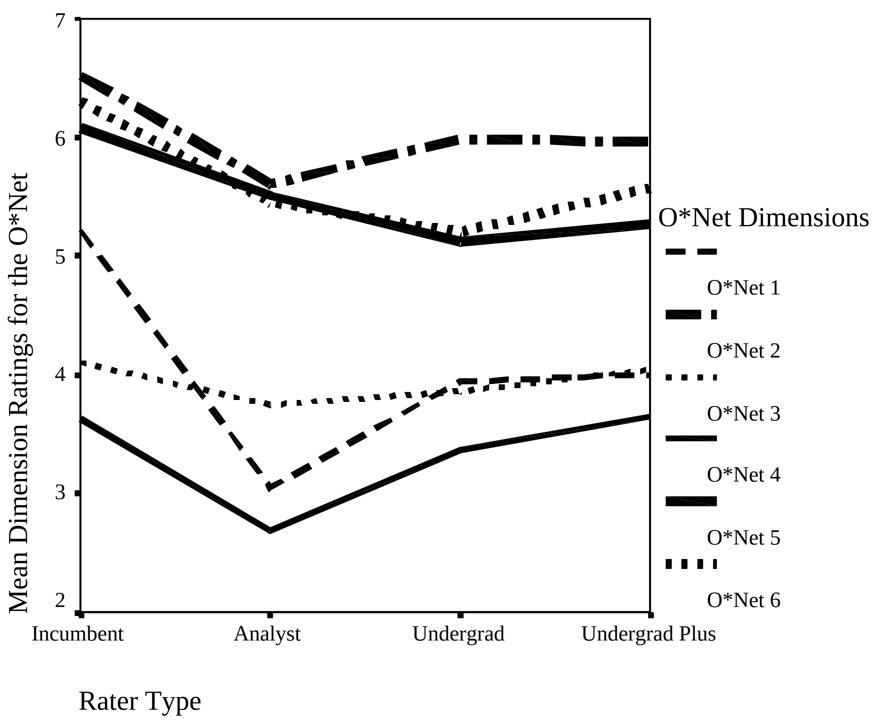
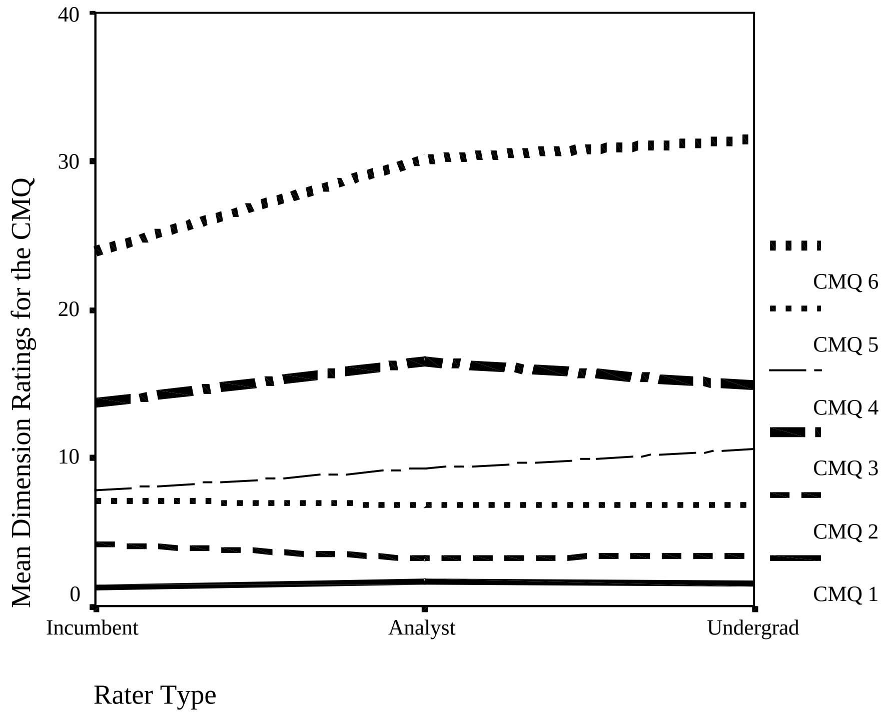


Figure 2. Mean Dimension Ratings for the CMQ by Rater



Discussion

The overarching aim of this study was to consider the sagacity of utilizing differing types of job analysis instruments (decomposed and holistic) and differing types of raters (incumbents, analysts, and job-naïve) for the purpose of obtaining information that describes a job of interest. Of greatest concern was the potential for one type of instrument or rater to provide ratings that were more or less accurate than others. The results obtained here help to shed light on this matter.

Although not direct measures of job analysis accuracy, measures of interrater agreement provide important insight into the degree to which raters agree among themselves regarding the work performed on a job. Independent of the type of rating instrument utilized, interrater agreements were highest for job analysts in this study. Reviewing the social psychological literature previously presented, there are several possible reasons that this might occur – job analysts may have highly developed stereotypes for the position of secretary, and may rely on this stereotype very heavily in the absence of more direct job information, job analysts should not suffer from the pressures of self-presentation that actual incumbents may feel, and finally, it is possible that lack of direct job knowledge prevents analysts from rating true between-job variability that might attenuate the agreement indices for incumbents. Interestingly, job-naïve raters often displayed more interrater agreement than did incumbents, as well. This might very well be attributable to the same processes that increase the interrater agreements among analysts.

It is also of interest to note if interrater agreements vary as a function of the type of instrument utilized. An instrument that provides a mechanism allowing for consistently stronger interrater agreements (without artificially inflating agreements with "Does Not Apply" ratings) would be ideal. Unfortunately, this study did not show the superiority of either holistic or decomposed rating scales in this regard. Interrater agreements between the instruments varied by rater type and whether or not indices were computed at the scale vs. dimension level. No discernable pattern of high or low interrater agreements is found in the data. Evidence did suggest that for scales comprised of both importance and level items, some gains in agreement could be obtained by looking at the variables separately. With regard to the O*Net, when considering all types of raters there appears to be some tendency for greater degrees of interrater agreement on the Importance scales as opposed to the Level scales. This might reflect the fact that many secretaries, while considering similar behaviors to be important, may actually need to perform these behaviors at differing levels for their respective positions. From this it is concluded that if interrater agreement is the outcome variable of interest the wiser investment is in choosing "who" rates, not on "what".

Overall, none of the interrater agreement indices obtained were exceptionally strong nor confidence inspiring if one hopes to utilize job analysis information for important personnel functions. It was anticipated that the non-threatening manner in which this data was collected, combined with the fairly homogenous population of respondents would yield considerably better results. However, an informal examination of the responses would certainly cause one to question if accurate responses can ever be obtained. The following represent some of the more obvious incongruities:

For the O*Net~

- 9% of job naïve raters stated that secretaries do *not* process information (luckily 100% of the incumbents and analysts disagreed!)
- 18% of analysts, 17% of naïve raters, and 9% of incumbents stated that secretaries do *not* solve problems
- 29% of incumbents, 27% of analysts, and 26% of naïve raters stated that secretaries do *not* move or handle objects (even when given the example of changing the settings on a copy machine)
- 27% of incumbents, 12% of analysts, and 25% of naïve raters stated that secretaries *do* run, navigate, or maneuver equipment such as forklifts, air craft, and watercraft
- 9% of incumbents and analysts and 12% of naïve raters stated that secretaries do *not* document information (written, verbal, or electronic)
- 63% of incumbents, 20% of analysts, and 48% of naïve raters stated that secretaries sell or influence others
- 30% of naïve raters stated that secretaries do *not* perform administrative activities (again, 100% of analysts and incumbents knew otherwise!)

For the CMQ and Criterion Data Sets~

- 26% of incumbents, 47% of analysts and 48% of naïve raters stated that secretaries supervise upper level employees – *none* of the raters in the criterion data set endorsed this item
- 6% of incumbents, 7% of analysts, and 4% of naïve raters stated that secretaries engage in selling or persuading behaviors (contrasted with above!) – 100% of the criterion group reported *never* using knowledge or skill in business administration/sales and marketing
- 23% of incumbents, 9% of analysts, and 16% of naïve raters stated that secretaries do *not* use or operate office machines or equipment – 100% of the criterion data set reported that they *do* use or operate office machines and equipment
- 14% of incumbents, 9% of analysts, and 12% of naïve raters reported that secretaries *do* not rely on any sensory input from the environment – 7% of raters in the criterion data set reported likewise
- 6% of incumbents, 3% of analysts, and 5% of naïve raters reported that secretaries do *not* engage in either numerical or verbal activities – all members of the criterion data set reported engaging in at least one of the two activities
- 3% of naïve raters stated that secretaries provide therapy and/or treatment – 100% of incumbents, analysts, and those in the criterion data set rated this as “Does-Not-Apply”

- 29% of incumbents, 9% of analysts, and 23% of naïve raters reported that secretaries do *not* exchange information with people in their organizations – 9% of respondents in the criterion data set concurred with this

Although providing a bit of humor when read, when considered in the context of job analysis as a serious and integral aspect of organizational development, the above are highly disconcerting. Although it is perhaps understandable that job naïve raters would provide odd responses, particularly when they are undergraduate students, it was not expected that incumbents and analysts would provide ratings that run blatantly counter to what one "knows" that secretaries do. It should be pointed out; however, that many of the job-naïve raters that are included in the above overview were actually removed prior to conducting the analyses reported herein. This certainly supports the contention of Harvey and Wilson (2000) that all job analysis ratings should be subjected to review and not taken as accurate at face value. In terms of the current study, it is reaffirmed that not only are the levels of interrater agreement obtained here inadequate, but also potentially misleading.

Whereas interrater agreement is important, criterion related validity is a more direct means of assessing the accuracy of job analysis ratings. The goal here was to determine the degree to which ratings obtained by different types of raters converged with an external job analysis conducted on the same position and to determine what type of instrument provided ratings that better converged with this external job analysis. The results here were quite clear, the decomposed measure was strongly and significantly correlated with the criterion data set, the holistic measure was not. Furthermore, with the decomposed measure, all types of raters showed strong convergence with the criterion (although analysts had the highest correlation, it was not significantly different from the correlations other raters had with the criterion); with the holistic measure, none did. This finding is consistent with previous findings in the job analysis field (Butler & Harvey, 1988, Sanchez & Levine, 1994). While several such studies have hypothesized that these results stem from the difficulty individuals have in making assessments that are cognitively complex in nature, none have employed manipulations that allow for examining this in a method consistent with cognitive psychology's conceptualizations of complexity.

It has typically been taken for granted in the job analysis field that decomposed ratings are cognitively simple to make and holistic ratings are complex. In cognitive psychology complexity is positively correlated with the number of attributes that must be considered or the number of judgments that must be made simultaneously (Morera & Budescu, 1998). It is not enough to assume that when making a holistic judgment of some attribute (e.g. communicating with others) that individuals simultaneously consider all the different aspects of this attribute (e.g. types of people they can possibly interact with, in all the possible situations they may encounter, and all the possible reasons for communicating). By creating a holistic condition in which participants are actually presented with the numerous attributes that would be encountered in a decomposed measure, but are required to consider them concurrently as opposed to individually, a truly complex decision making process can be better ensured. This study manipulated complexity in this manner in the hope of shedding some light on the repeated finding that decomposed job analysis measures are more accurate. Consistent with much of the decision making literature, and supporting the hypotheses made herein, individuals who made job ratings on the cognitively complex measure had correlations that most differed with the criterion data set (although it is noted that while in the predicted direction, this correlation was not significantly different from the other job naïve raters).

The implications of manipulating cognitive complexity were also not clear-cut with regard to interrater agreement, either. At the scale level, agreement indices were nominally lower for the cognitively complex manipulation than for those utilizing either the plain holistic instrument or the decomposed instrument. However, within the six dimensions there was greater variability and no clear trend emerged. Although conjecture, it is hypothesized that while making the rating task itself more difficult, the listing of attributes provided a common framework for raters to consider, thus resulting in greater levels of agreement. This would be consistent with findings that indicate that as the number of elements to be combined increases, the ratings of said attributes become more extreme - and therefore more homogenous (Manis et al., 1966). This would inflate interrater agreements found in highly complex decisions thru aggregation bias. The findings here are also consistent with studies of knowledge retrieval that have found that the consistency with which subjects make ratings is increased as the level of structure provided is increased (MacGregor et al., 1988). By providing subjects with a list of work behaviors, a larger degree of structure was imposed on the rating task, and respondents were essentially aided in recalling appropriate behaviors to go with each holistic dimension.

In addition to considering interrater agreements and accuracy, this study indicates that there may be strong tendencies of incumbents to provide considerably higher ratings of job behaviors than do other types of raters. Although there were only a few statistically significant mean differences by rater type, visual examination of Figure 1 shows that for all dimensions, incumbents always provided the highest ratings. This trend was not nearly as pronounced for the decomposed instrument. It is theorized that the holistic instrument by nature allows for stronger self-presentation strategies to be exhibited as it is more difficult to objectively verify the ratings provided and true quantification of ratings is not required.

A final issue examined in this study was the degree to which holistic and decomposed measures converged with one another. Although this does not in any way imply accuracy of one instrument or another (Slovic et al., 1977 in Morera & Budescu, 1998), it does help to indicate if instruments that appear to measure the same behaviors actually do. The very low correlations obtained between the holistic and decomposed measures utilized here could imply three things – a lack of construct validity on the part of one or both, that some non-random factor or circumstance intervened in my findings, or that both instruments actually measure the same things, however it was “missed” by how the dimensions were matched up. The third alternative appears highly unlikely in light of the Multiple Rs obtained in the regression analyses; it would seem difficult to argue that all the “important variance” is in there somewhere based on these exceedingly low and statistically insignificant values. The second alternative is dismissed based on the fact that the exact same participants provided ratings utilizing both measures, in a similar medium, within a very short time period. Consideration of some of the "crazy" ratings discussed previously might lead one to accept the first alternative based upon the premise of "garbage in – garbage out." However, the fact that there are low correlations between dimensions *across* the two instruments of interest does not necessarily imply this; there were significant correlations between dimensions *within* each instrument. If the ratings were all simply "random noise" or "bad" we wouldn't see this. The first alternative cannot be fully addressed with the current data; however it is noted that greater degrees of convergence were found for the decomposed instrument with the criterion job analysis; comparisons with other external criteria might provide additional insight.

Consistent with this, efforts were made to assess the degree to which the obtained ratings converged with a single "expert's" ratings utilizing both job analysis instruments. Appendix G reports findings of T tests that compare the dimension scores obtained here with the researcher's own ratings on the instruments of interest. Ironically, the researcher's dimension scores were significantly different from all other dimension scores – holistic ratings, decomposed ratings, and the criterion data set. When compared with earlier tests for mean differences between dimensions, the researcher did not perform favorably. Because there was no a clear cut method for examining interrater agreement for an N of 1, dimension scores on both the holistic and decomposed job analysis instruments were converted to percentiles as a means of comparing the researcher's ratings across the two instruments. The ratings are highly dissimilar. This, while disappointing, is not completely surprising in light of findings reported earlier – as there appears to be little convergence between the two instruments. From this (and consistent with the perspective that using multiple raters of any trait is more likely to yield accurate information) it is concluded that one should be cautious in utilizing any single individual's job analysis ratings as a gold standard.

As a further extension of making comparisons to other external criteria, future studies might build upon this one by utilizing a different type of criterion – one that is holistic in nature as opposed to decomposed. Although prior studies have utilized and recommended that decomposed instruments serve as the criterion against which other methodologies are compared, this might create an unfair advantage when testing decomposed measures. Because most decomposed instruments report dimension level data by mathematically aggregating the more molecular ratings, having raters simply rate at the dimension level is a viable alternative and would dispel any potential claims of bias in choosing a criterion.

In conclusion, the results of this study may serve as a warning to job analysis practitioners regarding the accuracy of judgments obtained utilizing holistic measures of work behavior. Whereas it appears that most raters can provide reasonably accurate ratings when using a decomposed instrument, the same cannot be said of holistic instruments. It is also noted that for holistic instruments, it would be a worthwhile investment of time and resources to structure the activity so that participants are less inclined to respond in a manner that exaggerates their ratings. Finally, it is noted that this study helps to provide support for the occurrence of cognitive psychological phenomena in the job analysis domain. It is a baby step in the direction proposed by Morgeson and Campion (1997) to integrate greater levels of theory into what many have considered to be an a-theoretical endeavor.

References

Anderson, L.E. (1999). Occupational preparation. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 91-104). Washington, DC: American Psychological Association.

Arad, S., Hanson, M.A., & Schneider, R.J. (1999). Organizational context. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 147-174). Washington, DC: American Psychological Association.

Armstrong, J.S., Denniston, W.B., & Gordon, M.M. (1975). The use of decomposition principle in making judgments. *Organizational Behavior and Human Decision Processes, 14*, 257-263.

Arvey, R.D. & Begalla, M.E. (1975). Analyzing the homemaker job using the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology, 60*, 513-517.

Ash, R.A. & Edgell, S.L. (1975). A note on the readability of the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology, 60*, 765-766.

Becker, B.E., & Cardy, R.L. (1986). Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration. *Journal of Applied Psychology, 71*, 662-671.

Benedict, M.E. & Levine, E.L. (1988). Delay and distortion: Tacit influences on performance appraisal effectiveness. *Journal of Applied Psychology, 73*, 507-514.

Bernadin, H.J. & Beatty, R.W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent.

Bethell-Fox, C.E. & Shepard, R.N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 12-23.

Borgen, F.H., Weiss, D.J., Tinsley, H.E.A., Dawis, R.V., & Lofquist, L.H. (1968). The measurement of occupational reinforcer patterns. *Minnesota Studies in Vocational Rehabilitation, XXV*. Minneapolis: University of Minnesota

Borman, W.C. (1975). Effects of instructions to avoid error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology, 60*, 556-56.

Borman, W.C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance, 20,* 238-252.

Borman, W.C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology, 64,* 410-421.

Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.

Borman, W.C., Kubisiak, U.C., & Schneider, R.J. (1999). Work styles. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 213-226). Washington, DC: American Psychological Association.

Butler, S.K., & Harvey, R.J. (1989). A comparison of holistic versus decomposed task rating of Position Analysis Questionnaire work dimensions. *Personnel Psychology, 41(4),* 761-771.

Carroll, J.B. (1993). Test theory and the behavioral scaling of test performance. In N. Frederiksen, R.J. Mislevy, & I.I. Behar (Eds.), *Test theory for a new generation of tests* (pp. 297-322). Hillsdale, NJ: Erlbaum.

Cooper, M.A., Schemmer, F.M., Jennings, M., & Korotkin, A.L. (1983*). Developing selection standards for Federal Bureau of Investigation special agents*. Bethesda, MD: Advanced Research Resources Organization.

Cooper, W.H. (1981). Ubiquitous halo. *Psychological Bulletin, 90,* 218-244.

Cornelius, E.T., DeNisi, A.S., & Blencoe, A.G. (1984). Expert and naïve raters using the PAQ: Does it matter? *Personnel Psychology, 37,* 453-464.

Cornelius, E.T., & Hakel, M.D. (1978). *A study to develop an improved enlisted performance evaluation system for the U.S. Coast Guard*. Washington, DC: Department of Transportation, United States Coast Guard.

Cornelius, E.T. & Lyness, K.S. (1981). A comparison of holistic and decomposed judgment strategies in job analyses by job incumbents. *Journal of Applied Psychology, 65(2),* 155-163.

Cornelius, E.T., Schmidt, F.L., & Carron, T.J. (1984). Job classification approaches and the implementation of validity generalization results. *Personnel Psychology, 37*, 247-26.

Costanza, D.P., Fleishman, E.A., & Marshall-Mies, J. (1999). Knowledges. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 71-90). Washington, DC: American Psychological Association.

Dawes, R.M. (1971). A case study of graduate admissions: Application of three principles of human decision-making. *American Psychologist, 26*, 180-188.

Dawis, R.V. (1991). Vocational interests, values, and preferences. In M.D. Dunnette and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 833-872). Palo Alto, CA: Consulting Psychologists Press.

DeNisi, A.S., Cornelius, E.T., III, & Blencoe, A.G. (1987). Further investigation of common knowledge effects on job analysis ratings. *Journal of Applied Psychology*, *72*, 262-268.

Driver, M.J. & Streufert, S. (1969). Integrative complexity: An approach to individuals and groups as information processing systems. *Administrative Science Quarterly, 14*, 272-285.

Einhorn, H.J. (1971). Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Decision Processes*, *6,* 1-27.

Einhorn, H.J. (1972). Expert measurement and mechanical combination. *Organizational Behavior and Human Decision Processes, 7,* 86-106.

Fine, S.A., & Wiley, W.W. (1971). An introduction to functional job analysis: Methods for manpower analysis (Monograph 4). Kalamazoo, MI: W.E. Upjohn Institute. Rand McNally.

Fischer, G. W. (1976). Multidimensional utility models for risky and reckless decisions. *Organizational Behavior and Human Performance, 17,* 127-146.

Fleishman, E.A. (1975). Toward a taxonomy of human performance. *American Psychologist, 30*, 1127-1149.

Fleishman, E.A. (1992). *Rating scale booklet: F-JAS*. Palo Alto, CA: Consulting Psychologists Press.

Fleishman, E.A. & Mumford, M. (1991). Evaluating classifications of job behavior: A construct validation of the ability requirement scales. *Personnel Psychology, 44*, 253-575.

Fleishman, E.A. & Reilly, M.E. (1992). *Administrator's Guide: F-JAS*. Palo Alto, CA: Consulting Psychologists Press.

Fleishman, E.A., Costanza, D.P., & Marshall-Mies, J. (1999). Abilities. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 175-198). Washington, DC: American Psychological Association.

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

Friedman, L., & Harvey, R.J. (1986). Can raters with reduced job descriptive information provide accurate Position Analysis Questionnaire (PAQ) ratings? *Personnel Psychology, 39,* 779-789.

Gael, S. (1983). *Job analysis: A guide to assessing work activities*. San Francisco: Jossey-Bass.

Goldberg, L.R. (1967). Research on the clinical judgment process. *Gawein*, 15, 220-242.

Goldberg, L.R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23*, 483-496.

Green, S.B. & Veres, J.G. (1990). Evaluation of an index to detect inaccurate respondents to a task analysis inventory. *Journal of Business and Psychology, 5*, 47-61.

Hahn, D.C. & Dipboye, R.L. (1988). Effects of training and information on the accuracy and reliability of job evaluations. *Journal of Applied Psychology, 73*, 146-153.

Hakel, D.T. (1986). Personnel selection and placement. *Annual Review of Psychology, 37*, 351-38.

Harris, M.M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.

Harvey, R.J. (1991). Job Analysis. In M.D. Dunnette and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 71-163). Palo Alto, CA: Consulting Psychologists Press.

Harvey, R.J. (1993). *Research monograph: Development of the Common-Metric Questionnaire (CMQ).* Personnel Systems and Technologies Corporation & Virginia Polytechnic Institute & State University.

Harvey, R.J., Friedman, L., Hakel, M.D., & Cornelius, E.T., (1988). Dimensionality of the Job Element Inventory, a simplified worker-oriented job analysis questionnaire. *Journal of Applied Psychology, 73*, 639-646.

Harvey, R.J. & Hayes, T.L. (1986). Monte Carlo baselines for interrater reliability correlations using the Position Analysis Questionnaire. *Personnel Psychology, 39*, 345-357.

Harvey, R.J., & Lozada-Larsen, S.R. (1988). Influence of amount of job descriptive information on job analysis rating accuracy. *Journal of Applied Psychology, 73*, 457-461.

Harvey, R.J. & Wilson, M.A. (2000). Yes Virginia, there is an objective reality I job analysis. *Journal of Organizational Behavior, 21*, 829-854.

Hogan, J.C., Ogden, G.D., & Fleishman, E.A. (1978). *Assessing the physical requirements in selected benchmark jobs* (Final Rep. 3012). Bethesda, MD: Advanced Research Resources Organization.

Hogan, J.C., Ogden, G.D., & Fleishman, E.A. (1979). *The development and validation of tests for the order selector job at Certified Grocers of California, Ltd*. Bethesda, MD: Advanced Research Resources Organization.

Hora, S.C., Dodd, N.G., & Hora, J.A. (1993). The use of decomposition in probability assessments of continuous variables. *Journal of Behavioral Decision Making, 6*, 133-147.

Jacoby, J., Kuss, A., Mazursky, D., & Troutman, T. (1985). Effectiveness of security analyst information accessing strategies: A computer interactive assessment. *Computers in Human Behavior, 1*, 95-113.

Jacoby, J., Speller, D.E., & Berning, C.K. (1974). Brand choice behavior as a function of information load: Replication and extension. *Journal of Consumer Research, 1*, 33-42.

Jacoby, J., Speller, D.E., & Kohn, C.A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research, 11*, 63-69.

Jeanneret, P.R., Borman, W.J., Kubisiak, U.C., & Hanson, M.A. (1999). Generalized work activities. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 105-125). Washington, DC: American Psychological Association.

Jones, A.P., Main, D.S., Butler, M.C., & Johnson, L.A. (1982). Narrative job descriptions as potential sources of job analysis ratings. *Personnel Psychology, 35*, 813-828.

Keeley, S.M. & Doherty, M.E. (1972). Bayesian and regression modeling of graduate admission policy. *Organizational Behavior and Decision Processes, 8*, 297-323.

Kozlowski, S.W.J., Kirsch, M.P., & Chao, G.T. (1986). Job knowledge, ratee familiarity, conceptual similarity, and halo error: An exploration. *Journal of Applied Psychology, 71*, 45-49.

Landy, F.J. (1988). Selection procedure development and usage. In S. Gael (Ed.), *The job analysis handbook for business, government, and industry* (pp. 271-287). New York: Wiley.

Lyness, K.S. & Cornelius, E.T. (1982). A comparison of holistic and decomposed judgment strategies in a performance rating simulation. *Organizational Behavior and Human Decision Processes, 29,* 21-38.

Mabe, P.A. & West, S.G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67,* 280-296.

MacGregor, D., Lichtenstein, S., & Slovic, P. (1988). Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. *Organizational Behavior and Human Decision Processes, 4,* 303-323.

Manis, M., Gleason, T.C., & Dawes, R.M. (1966). The evaluation of complex social stimuli. *Journal of Personality and Social Psychology, 3,* 404-419.

Marlowe, D. & Crowne, D.P. (1961). Social desirability and response to perceived situational demands. *Journal of Consulting Psychology, 25,* 109-115.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Morera, O.F. & Budescu, D.V. (1998). A psychometric analysis of the "Divide and Conquer" principle in multicriteria decision making. *Organizational Behavior and Human Decision Processes, 75,* 187-206.

Morgeson, F.P. & Campion, M.A. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology, 82*, 627-655.

Morgeson, F.P. & Campion, M.A. (2000). Accuracy in job analysis: toward an inference-based model. *Journal of Organizational Behavior*, *21,* 819-827.

Mumford, M.D., Peterson, N.G., & Childs, R.A. (1999). Basic and cross-functional skills. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 49-69). Washington, DC: American Psychological Association.

Mumford, M.D., Yarkin-Levin, K., Korotkin, A.C., Wallis, M.R., & Marshall-Mies, J. (1985). *Characteristics relevant to performance as an Army leader: Knowledge, skills, abilities, and other characteristics, and general skills* (Tech Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Murphy, K.R., Garcia, M., Kerkar, S., Martin, C., & Balzar, W.K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. *Journal of Applied Psychology, 67,* 320-325.

Nisbett, R. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65*, 373-406.

Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A., (1999). *An occupational information system for the 21st century: The development of the O*Net*. Washington, DC: American Psychological Association.

Prien, E.P. & Saleh, S.D. (1963). A study of bias in job analysis and evaluation. *Journal of Industrial Psychology, 1*, 113-117.

Ravinder, R.H., Kleinmuntz, D.N. (1991). Random error in additive decompositions of multiattribute utility. *Journal of Behavioral Decision Making, 4,* 83- 97 .

Roberts, K.H. & Glick, W. (1981). The job characteristics approach to task design: A critical review. *Journal of Applied Psychology, 66*, 193-217.

Sackett, P.R., Cornelius, E.T., & Carron, T.J. (1981). A comparison of global judgment vs. ask oriented approaches to job classification. *Personnel Psychology, 34*, 791-804.

Sager, C.E. (1999). Occupational interests and values. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 197-211). Washington, DC: American Psychological Association.

Sanchez, J.I. & Fraser, S.L. (1992). On the choice of scales for task analysis. *Journal of Applied Psychology, 77*, 545-553.

Sanchez, J.I. & Levine, E.L. (1994). The impact of raters' cognition on judgment accuracy: An extension to the job analysis domain. *Journal of Business and Psychology, 9*, 47-57.

Sanchez, J.I. & Levine, E.L. (2000). Accuracy or consequential validity: which is the better standard for job analysis data? *Journal of Organizational Behavior, 21,*809-818.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178-20.

Schlenker, B.R. (1990). *Impression management: The self-concept, social identity, and interpersonal relations*. Monterey, CA: Brooks/Cole.

Smith, B.N., Hornsby, J.S., Benson, P.G., & Wesolowski, M. (1989). What is in a name: The impact of job titles on job evaluation results. *Journal of Business and Psychology, 3*, 341-351.

Smith, J.E., & Hakel, M.D. (1979). Convergence among data sources, response bias, and reliability and validity of a structured job analysis questionnaire. *Personnel Psychology, 32,* 677-692.

Society for Industrial and Organizational Psychology, Inc. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.) College Park, MD: Author.

Streufert, S., Suedfeld, P., & Driver, M.J. (1965). Conceptual structure, information search, and information utilization. *Journal of Personality and Social Psychology, 2*, 736-74.

Strong, M.H., Jeanneret, P.R., McPhail, S.M., Blakley, B.R., & D'Egidio, E.L. (1999). Work context: Taxonomy and measurement of the work environment. In Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of the O*Net* (pp. 127-145). Washington, DC: American Psychological Association.

Thorndike, E.L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 25-29.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Uniform guidelines on employee selection procedures (1978). *Federal Register*, 43, 38290-38315

U.S. Department of Labor. (1973). *Task analysis inventories*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. (1991). *Dictionary of occupational titles* (4th ed., rev.). Washington, DC: U.S. Government Printing Office.

Wilson, M.A., Harvey, R.J., & Macy, B.A. (1990). Repeating items to estimate the test-retest reliability of task inventory ratings. *Journal of Applied Psychology*, *75*, 158-163.

Wyer, R.S. & Srull, T.K. (1980). The processing of social stimulus information. A conceptual integration. In R. Hastie, T.M. Ostrom, E.B. Ebbesen, R.S. Wyer, D.L. Hamilton, & D.E. Carlston (Eds.), *Person memory: The cognitive basis of social perception* (pp. 227-300). Hillsdale, NJ: Erlbaum.

Appendix A

Job Description for “Secretary”

Job Description for Secretary

Performs routine clerical and administrative functions such as drafting correspondence, scheduling appointments, organizing and maintaining paper and electronic files, or providing information to callers.

Representative Tasks*

- Answers telephone and gives information to callers, takes messages, or transfers calls to appropriate individuals.
- Opens incoming mail and routes mail to appropriate individuals.
- Answers routine correspondence.
- Composes and distributes meeting notes, correspondence, and reports.
- Schedules appointments.
- Maintains calendar and coordinates conferences and meetings.
- Takes dictation in shorthand or by machine and transcribes information.
- Locates and attaches appropriate file to incoming correspondence requiring reply.
- Files correspondence and other records.
- Makes copies of correspondence and other printed matter.
- Arranges travel schedules and reservations.
- Greets and welcomes visitors, determines nature of business, and conducts visitors to employer or appropriate person.
- Compiles and maintains lists and records, using typewriter or computer.
- Records and types minutes of meetings, using typewriter or computer.
- Compiles and types statistical reports, using typewriter or computer.
- Mails newsletters, promotional material and other information.
- Orders and dispenses supplies.
- Prepares and mails checks.
- Collects and disburses funds from cash account and keeps records.
- Provides customer services such as, order placement and account information.

*This list of tasks is intended to be representative of the position and does not include all tasks performed by incumbents.

Appendix B

O*Net Generalized Work Activities Questionnaire

Appendix C

CMQ Questionnaire

Appendix D

Matched Dimensions for the CMQ and the O*Net

Dimension Name	CMQ Dimensions Comprising This	O*Net Dimensions Comprising This
D1: Training, Managing, and Coordinating Others	1 – 9. Supervising & consulting internal contacts (IC) 11. training IC 20 – 21. Supervising external contacts (EC) 25. training EC	8. Coordinating, developing, managing, and advising others
D2: Communicating and Interacting with Others	10. Selling IC 12 – 19. IC & EC Selling, entertaining, bargaining, resolving conflicts, etc. 26 – 39. Chairing and Attending meetings with various constituents	7. Communicating and interacting
D3: Working with Technical Equipment and Performing Technical Activities	63. Office machines and equipment (using, repairing, calibrating, etc.) 64. Technical and scientific equipment	6. Performing complex/technical activities
D4: Physical and Manual Work Activities	75. Gross Physical Activity 76. Precise Physical Activity	5. Performing physical and manual work activities
D5: Decision Making with Regard to Company Resources	48. Production/Operations mgmt. 49. Human resource mgmt. 50 – 51. Financial decisions 52 – 53. Strategic Planning	4. Reasoning/decisions making 9. Administering
D6: Obtaining, Evaluating, or Processing Information from the Work Environment	77. Sensory input from the work environment 78. Numerical and verbal activities	1. Looking for and receiving job related information 2. Identifying and evaluating job-related information 3. Information/data processing

Appendix E

Test for Differences Between Correlations In Total Job Analysis Scores By Rater Type

Tests for Correlation Differences by Rater Type on the O*Net

Test 1: O*Net Incumbent Correlations < O*Net Analyst Correlations
$t = .453$
($df = 67$)

Test 2: O*Net Incumbent Correlations > O*Net Job Naïve Correlations
$t = 1.067$
($df = 93$)

Test 3: O*Net Incumbent Correlations > O*Net Job Naïve Complex Correlations
$t = 1.365$*
($df = 93$)

Test 4: O*Net Analyst Correlations > O*Net Job Naïve Correlations
$t = 1.612$
($df = 92$)

Test 5: O*Net Analyst Correlations > O*Net Job Naïve Complex Correlations
$t = 2.112$*
($df = 92$)

Test 6: O*Net Job Naïve Correlations > O*Net Job Naïve Complex Correlations
$t = .4583$
($df = 118$)

Tests for Correlation Differences by Rater Type on the CMQ

Test 1: CMQ Incumbent Correlations < CMQ Analyst Correlations
$t = .048$
($df = 67$)

Test 2: CMQ Incumbent Correlations > CMQ Job Naïve Correlations
$t = .0860$
($df = 153$)

Test 3: CMQ Analyst Correlations > CMQ Job Naïve Correlations
$t = .0911$
($df = .152$)

*significant at $p = .05$.

Appendix F

Correlations by Rater Type For Each Dimension on Each Scale

Correlations by Rater Type - Dimension 1.

	O*Net D1 Incumbent	O*Net D1 Analyst	O*Net D1 Job Naive	O*Net D1 Job Naive Complex	CMQ D1 Incumbent	CMQ D1 Analyst	CMQ D1 Job Naive	Criterion D1
O*Net D1 Incumbent	1.000	-.054	.010	.101	-.095	.248	.154	-.237
O*Net D1 Analyst		1.000	-.113	-.137	-.063	-.003	.095	.314
O*Net D1 Job Naive			1.000	.016	-.080	.005	.099	-.062
O*Net D1 Job Naive Complex				1.000	-.088	-.128	.216	-.276
CMQ D1 Incumbent					1.000	-.014	-.135	.495*
CMQ D1 Analyst						1.000	.118	.792*
CMQ D1 Job Naive							1.000	.491*
Criterion D1								1.000

* Correlation is significant at the .05 level (2-tailed).

Correlations by Rater Type - Dimension 2.

	O*Net D2 Incumbent	O*Net D2 Analyst	O*Net D2 Job Naive	O*Net D2 Job Naive Complex	CMQ D2 Incumbent	CMQ D2 Analyst	CMQ D2 Job Naive	Criterion D2
O*Net D2 Incumbent	1.000	.190	-.009	-.010	-.138	-.106	.147	-.325
O*Net D2 Analyst		1.000	.245	.142	-.121	-.258	.170	.241
O*Net D2 Job Naive			1.000	.018	.075	-.218	-.152	-.172
O*Net D2 Job Naive Complex				1.000	.011	-.081	.141	-.379
CMQ D2 Incumbent					1.000	-.111	-.065	.731*
CMQ D2 Analyst						1.000	.067	.654*
CMQ D2 Job Naive							1.000	.369
Criterion D2								1.000

* Correlation is significant at the .05 level (2-tailed).

Correlations by Rater Type - Dimension 3.

	O*Net D3 Incumbent	O*Net D3 Analyst	O*Net D3 Job Naive	O*Net D3 Job Naive Complex	CMQ D3 Incumbent	CMQ D3 Analyst	CMQ D3 Job Naive	Criterion D3
O*Net D3 Incumbent	1.000	-.385*	.051	.038	.071	.042	-.088	.181
O*Net D3 Analyst		1.000	-.096	-.219	-.056	-.068	-.098	.235
O*Net D3 Job Naive			1.000	-.013	.147	.178	-.065	-.177
O*Net D3 Job Naïve Complex				1.000	.142	.047	-.150	-.258
CMQ D3 Incumbent					1.000	-.016	.212	.424*
CMQ D3 Analyst						1.000	.281	.624*
CMQ D3 Job Naive							1.000	.587*
Criterion D3								1.000

* Correlation is significant at the .05 level (2-tailed).

Correlations by Rater Type - Dimension 4.

	O*Net D4 Incumbent	O*Net D4 Analyst	O*Net D4 Job Naive	O*Net D4 Job Naive Complex	CMQ D4 Incumbent	CMQ D4 Analyst	CMQ D4 Job Naive	Criterion D4
O*Net D4 Incumbent	1.000	-.075	-.027	.017	-.117	-.383*	.039	-.113
O*Net D4 Analyst		1.000	.171	-.147	.122	.027	.281	.189
O*Net D4 Job Naive			1.000	-.011	-.053	.077	-.077	-.148
O*Net D4 Job Naive Complex				1.000	-.056	.084	-.260*	.121
CMQ D4 Incumbent					1.000	-.027	-.101	.631*
CMQ D4 Analyst						1.000	-.074	.887*
CMQ D4 Job Naive							1.000	.277
Criterion D4								1.000

* Correlation is significant at the .05 level (2-tailed).

Correlations by Rater Type - Dimension 5.

	O*Net D5 Incumbent	O*Net D5 Analyst	O*Net D5 Job Naive	O*Net D5 Job Naive Complex	CMQ D5 Incumbent	CMQ D5 Analyst	CMQ D5 Job Naive	Criterion D5
O*Net D5 Incumbent	1.000	-.050	-.204	.045	.116	-.179	-.107	.045
O*Net D5 Analyst		1.000	.078	.111	.124	-.358*	.246	.336
O*Net D5 Job Naive			1.000	-.019	-.181	.023	-.081	-.189
O*Net D5 Job Naive Complex				1.000	.124	.382*	-.068	-.296
CMQ D5 Incumbent					1.000	.022	-.065	.457*
CMQ D5 Analyst						1.000	-.042	.598*
CMQ D5 Job Naive							1.000	.393
Criterion D5								1.000

* Correlation is significant at the .05 level (2-tailed).

Correlations by Rater Type - Dimension 6

	O*Net D6 Incumbent	O*Net D6 Analyst	O*Net D6 Job Naive	O*Net D6 Job Naive Complex	CMQ D6 Incumbent	CMQ D6 Analyst	CMQ D6 Job Naive	Criterion D6
O*Net D6 Incumbent	1.000	-.087	-.401*	.014	.131	-.080	.258	-.283
O*Net D6 Analyst		1.000	-.027	-.277	.050	-.384*	.326	.227
O*Net D6 Job Naive			1.000	-.091	-.091	.103	.018	-.174
O*Net D6 Job Naive Complex				1.000	.119	.086	-.108	-.457*
CMQ D6 Incumbent					1.000	-.103	-.058	.756*
CMQ D6 Analyst						1.000	.041	.492*
CMQ D6 Job Naive							1.000	.586*
Criterion D6								1.000

* Correlation is significant at the .05 level (2-tailed).

Appendix G

Mean Dimension Scores
For All Scales
By Rater Type

Descriptive Statistics for CMQ

CMQ Rater Type		Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6
Incumbent	Mean	1.2484	4.1484	14.7143	8.6857	8.2048	23.9143
	Std. Deviation	2.9987	2.1995	13.6227	8.8174	7.8788	12.3773
Analyst	Mean	1.7596	3.7091	17.2656	9.8469	6.9406	3.1562
	Std. Deviation	2.4303	2.1787	16.0035	5.6851	4.3622	13.5572
Job Naive	Mean	1.4472	3.8206	13.5347	1.7970	6.4267	31.4505
	Std. Deviation	6.2049	5.5235	13.9351	9.2730	1.3619	15.0678

Descriptive Statistics for O*Net

O*Net Rater Type		Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6
Incumbent	Mean	5.1762	6.4714	4.1886	3.6810	6.0476	6.2889
	Std. Deviation	2.4128	1.7932	1.2461	2.2699	2.2934	1.4634
Analyst	Mean	3.3500	5.8103	3.9706	2.7471	5.8012	5.7999
	Std. Deviation	2.2260	1.7177	1.4100	1.6711	1.7818	1.9357
Job Naive	Mean	4.0944	6.2896	4.0100	3.6628	5.6306	5.6981
	Std. Deviation	2.2414	1.6204	1.3552	2.0473	1.6333	1.7208
Job Naive	Mean	4.0926	6.1042	4.0245	3.8028	5.6639	5.8685
Complex	Std. Deviation	2.1897	1.7362	1.4329	2.3810	1.7271	1.5236

Descriptive Statistics for Criterion Data

	Mean	Std. Deviation
Dimension 1	2.9508	2.1134
Dimension 2	4.6923	2.3860
Dimension 3	13.7300	7.4356
Dimension 4	1.4600	13.2460
Dimension 5	7.1367	4.1340
Dimension 6	26.8600	9.9078

Appendix H

Convergence of Collected Data With Researcher's Data & Convergence of Researcher's Data Across Job Analysis Instruments

Researcher's O*Net vs. Collected O*Net

Researcher's O*Net Dimension 1 Value = 5.83						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
O*Net D1	-10.992	188	.000	-1.7921	-2.1137	-1.4705

Researcher's O*Net Dimension 2 Value = 7.25						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
O*Net D2	-9.989	188	.000	-1.2401	-1.4850	-.9952

Researcher's O*Net Dimension 3 Value = 4.80						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
O*Net D3	-8.649	188	.000	-.8571	-1.0526	-.6616

Researcher's O*Net Dimension 4 Value = 6.83						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
O*Net D4	-5.453	188	.000	-.8529	-1.1615	-.5444

Researcher's O*Net Dimension 5 Value = 5.00						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
O*Net D5	3.133	188	.002	.4145	.1535	.6754

Researcher's O*Net Dimension 6 Value = 6.33						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
O*Net D6	-6.335	188	.000	-.7608	-.9977	-.5239

Researcher's CMQ vs Collected CMQ

Researcher's CMQ Dimension 1 Value = 15.31						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CMQ D1	-17.697	188	.000	-8.5018	-9.4503	-7.5534

Researcher's CMQ Dimension 2 Value = 15.04						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CMQ D2	-25.344	188	.000	-11.0771	-11.9400	-10.2142

Researcher's CMQ Dimension 3 Value = 65.5						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CMQ D3	-43.312	188	.000	-47.6161	-49.7865	-45.4456

Researcher's CMQ Dimension 4 Value = 41						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CMQ D4	-11.479	188	.000	-16.1667	-18.9472	-13.3861

Researcher's CMQ Dimension 5 Value = 9.67						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CMQ D5	-4.256	188	.000	-2.9458	-4.3123	-1.5793

Researcher's CMQ Dimension 6 Value = 37						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CMQ D6	-6.586	188	.000	-7.3661	-9.5742	-5.1579

<u>Researcher's CMQ vs. Criterion CMQ</u>

Researcher's CMQ Dimension 1 Value = 15.31						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CRD1	-8.794	49	.000	-6.3592	-7.8124	-4.9060

Researcher's CMQ Dimension 2 Value = 15.04						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CRD2	-30.665	49	.000	-10.3477	-11.0258	-9.6696

Researcher's CMQ Dimension 3 Value = 65.5						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CRD3	-39.722	49	.000	-41.7700	-43.8832	-39.6568

Researcher's CMQ Dimension 4 Value = 41						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CRD4	-3.491	49	.001	-6.5400	-10.3045	-2.7755

Researcher's CMQ Dimension 5 Value = 9.67						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CRD5	-9.464	49	.000	-5.5333	-6.7082	-4.3584

Researcher's CMQ Dimension 6 Value = 37						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
CRD6	-6.523	49	.000	-9.1400	-11.9558	-6.3242

Researcher's O*Net vs. Researcher's CMQ

Dimension	O*Net Percentage Score	CMQ Dimension Score
Dimension 1	.4171	.1531
Dimension 2	.5186	.1504
Dimension 3	.2714	.6550
Dimension 4	.3757	.4100
Dimension 5	.3400	.0967
Dimension 6	.4543	.3700

BLDG. 486, #210A WILSON DRIVE • HAVELOCK, NC 28532
PHONE (540) 320-8724 • E-MAIL SHANANGIBSON@HOTMAIL.COM

SHANAN GWALTNEY GIBSON

OBJECTIVE

Employment with an organization that allows for utilization of both my graduate level training and experience in the applied and research-oriented aspects of Industrial & Organizational Psychology, Human Resource Management, and Personnel Issues. Strong preference for diverse projects, internal and/or external consulting, and travel opportunities.

SUMMARY OF QUALIFICATIONS

- *Strong Project Management, Interpersonal, & Group Facilitation Skills*
- *Extensive Knowledge of Human Resources & Personnel Management*
  - Personnel Recruitment & Selection
  - Personnel Law
  - Training and Development
  - Performance Appraisal
  - Job Analysis
  - Organizational Development & Leadership
  - Compensation Theory
- *Highly Advanced Quantitative Skills*
  - Research Design & Methodology
  - Statistical Data Analysis – ANOVA, Multivariate Regression, Factor Analysis, Item Response Theory, Differential Item Functioning, Structural Equation Modeling
- *Computer Competencies*
  - Microsoft Office Suite
  - Statistical: SAS, SPSS, Bilog

EDUCATION

May 2001 — *Ph.D. Industrial/Organizational Psychology*
Virginia Polytechnic Institute and State University
Dissertation: "Call Me Old Fashioned – Is My Job Analysis Accurate or Not?"

Dec. 1998 — *M.S. Industrial/Organizational Psychology*
Virginia Polytechnic Institute and State University
Thesis: "Differential Item Functioning on the Armed Services Vocational Aptitude Battery (ASVAB)"

June 1995 — *B.A. General Studies – Magna Cum Laude*
Armstrong Atlantic State University

EMPLOYMENT

08/00 – 05/01 Radford, VA	*Radford University, Adjunct Professor of Psychology.* Fall Courses: Psychology of Work Behavior & Social Psychology Spring Courses: Research Methods and Statistics for Behavioral Science & Social Psychology
08/00 – 05/01 Blacksburg, VA	*VA Tech, Instructor of Psychology.* Fall Course: Advanced Social Psychology for Majors Spring Course: Social Psychology
08/99 – 08/00 Corpus Christi, TX	*Legacy of Life Tissue Foundation, Regional Manager.* Maintain public relations and contractual agreements with fourteen regional hospitals; facilitate relationships among hospital CEOs, medical staff, federally designated Organ Procurement Organizations, and technical staff. Assure compliance of hospital policies with relevant federal and state laws. Develop and provide public and professional education/training on organ and tissue donation and recruit potential donors in the Coastal Bend region of Texas.
01/99 – 05/00 Corpus Christi, TX	*TX A&M University, Adjunct Professor of Psychology.* Course: Industrial & Organizational Psychology
08/98 – 06/99 Blacksburg, VA	*Research Associate, VA Tech Dept. of Computer Science.* Work as part of a team to conduct research in an on-going study tracking the interaction of students with computers in the classroom, and the subsequent impact on learning. Create surveys; collect critical incident data; contribute to the creation of a framework for evaluation and assessment of outcomes; facilitate the integration of mentors into the project-based, computer-mediated, learning environment; and develop manuscripts for publication.
05/98 – 08/98 Washington, DC	*American Psychological Association (APA), Summer Intern.* Function as assistant in the Testing and Assessment division of the Science Directorate. Conduct research; track pending legislation and court cases; and write relevant articles and research synapses as they relate to EEO, workplace harassment and discrimination, hiring and selection.
08/96 – 05/98 Blacksburg, VA	*VA Tech Division of Continuing Education, Job Analyst and Research Associate.* Conduct job analyses at local organizations utilizing ACT's Work Keys job profiling system; research worker shortage and training issues in the high-tech/IT sector. Create and conduct surveys; write

tech/IT sector. Create and conduct surveys; write technical reports; serve as representative for the Division of Continuing Education at meetings and conferences

10/92 – 06/96
Savannah, GA

Critz, Inc., Customer Relations Representative & Sales Associate.
Develop & implement programs to assess & maintain positive customer relations; sales and leasing associate for Mercedes Benz and BMW automobiles

PROFESSIONAL MEMBERSHIPS

American Psychological Association (student affiliate)
Society of Industrial/Organizational Psychology (student affiliate)
Personnel Testing Council of Metropolitan Washington DC (student affiliate)
International Personnel Mgmt. Assoc. Assessment Council (student affiliate)

PUBLICATIONS

Gibson, S.G. & Harvey, R.J. (in press). Differential item functioning on the armed services vocational aptitude battery. *Journal of Educational and Psychological Measurement*.

Gibson, S., Neale, D.C., Carroll, J.M. and Van Metre, C.A. (1999). Telementoring in a school environment. *Proceedings of CSCL 99: Computer Supported Cooperative Learning.* pp. 182-188. Mahwah, NJ: Lawrence Erlbaum.

Gibson, S.G. & Fox, H.R. (Oct. 1998). Supreme court rulings in sexual harassment disputes. *The Industrial-Organizational Psychologist, 35*, 33-36.

Fox, H.R. & Gibson, S.G. (Aug. 1998). Supreme court rulings in sexual harassment disputes. *The Personnel Testing Council of Metropolitan Washington DC Newsletter*. Washington, DC.

Leffel, L., Frary, R., Gwaltney, S., & Schaefermeyer, M. (1998). *Help Wanted: A call for collaborative action for the new millennium* (Research Report). VA: Virginia Tech Division of Continuing Education in collaboration with the Information Technology Association of America (ITAA).

Leffel, L., Schaefermeyer, M., Smythson, J., & Gwaltney, S. (1997). *NICHE - Northern Virginia Initiative on Continuing Higher Education* (Technical Report). Blacksburg, VA: Virginia Tech Division of Continuing Education.

CONFERENCE PRESENTATIONS

Gibson, S.G. (Apr. 2001). *The dimensionality of the ASVAB: What do we really know?* Poster presented at the annual meeting of the Society for Industrial/Organizational Psychology, San Diego, CA.

Gibson, S.G., Neale, D., & Carroll, J. (Dec. 1999). *Tele-mentoring in a project-based learning environment.* Paper presented at the annual meeting of Computer Supported Collaborative Work, Stanford University.

Gibson, S.G. & Harvey, R.J. (Apr. 1999). *Differential item functioning on the armed services vocational aptitude battery.* Poster presented at the annual meeting of the Society for Industrial/Organizational Psychology, Atlanta, GA.

Gwaltney, S. (1998). *Survey Research and the Technical Skills Shortage. Focus on "Soft Skills"*. Closing the Gap, A National Forum on the Information Technology Skills Shortage. University Continuing Education Association, Washington, DC.

AWARDS

1998 UCEA Innovative Awards in Continuing Education. Leffel, L., Gwaltney, S., & Schaefermeyer, M. (1998). *NICHE - Northern Virginia Initiative on Continuing Higher Education*. National award presented annually to recognize the faculty and staff of University Continuing Education Association (UCEA) member institutions who are making innovative contributions to the field of higher education.

WORKS IN PROGRESS

Gibson, S.G. (2000*). Implicit leadership theories: Are all prototypes created equally?* Grant proposal under development.