BEHAVIORAL SPECIFICITY AND RELIABILITY IN JOB ANALYSIS AND JOB SPECIFICATION

by

Marc Cowgill

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

MASTER OF SCIENCE

in

Psychology

APPROVED:

R. J. Harvey, Chairperson

Roseanne Foti

Neil Hauenstein

December, 1991

Blacksburg, Virginia
BEHAVIORAL SPECIFICITY AND RELIABILITY IN JOB ANALYSIS AND JOB SPECIFICATION

by

Marc Cowgill

Committee Chairman: R. J. Harvey
Psychology

(ABSTRACT)

Job analysis, narrowly defined, refers to the collection of data describing job-related behaviors and the characteristics of the job environment. Job specification refers to the process of inferring required traits or abilities necessary for a desired level of job performance. Differences in the judgmental processes involved in these two functions were explored by (a) investigating the potential schema- or stereotype-based nature of job specification ratings, and (b) assessing the relationship between behavioral specificity and interrater reliability. These concerns were investigated through the use of 3 groups of subject raters: one group possessing extensive job knowledge, one group possessing some degree of job familiarity, and one group possessing little or no job knowledge. All subjects completed a job analysis instrument (the Job Element Inventory) and a job specification instrument (the Threshold Traits Analysis; TTA). Contrary to predictions, little evidence was uncovered to suggest extensive schema-usage on the part of TTA raters. In addition, the 2 instruments achieved similar levels of
interrater reliability among the 3 subject groups. However, marginal support was found for the notion that behaviorally specific items generate higher reliability the less-specific items, and in replication of previous findings, job-naive raters were found unable to achieve the reliability of subject matter experts. Suggestions for future research are offered.
ACKNOWLEDGEMENTS

I would like to thank my lab assistant, Fred Homan, for his generosity in volunteering his time and also for the professionalism of his labors. In addition, I would like to thank Professor R. J. Harvey. Without his expertise and direction, this study never could have been completed. I would also like to thank my other committee members, Professor Neil Hauenstein and Professor Roseanne Foti, for their guidance and support.

I am grateful for the help given to me by Dr. Gerald J. Kowalski, the staff of resident advisors, and the undergraduate student subjects.

Let me also thank the head of the psychology department, Dr. Joseph Sgro for his continual support.

Further acknowledgements will be reserved for more intimate modes of communication.
Table of Contents

General Section on Job Comparison Approaches .......... p. 1
General Section on Schema Research ..................... p. 6
Summary of Job Analysis Research to Date ................ p. 9
Considerations Regarding the Study ..................... p. 18
Hypotheses of the Study ................................ p. 24
Results of the Study ...................................... p. 29
General Discussion and Suggested Future Directions ... p. 36
References .................................................... p. 44

Appendix A (TTA) ........................................... p. 56
Appendix B (JEI) ............................................ p. 62
Appendix C (statistical formulae) ......................... p. 74
Vita ......................................................... p. 75
List of Figures

Figure 1: An Example of Behavior-Trait Linkages ...... p. 19

Figure 2: Rating Instrument as a function of IR reliability and Subject Group ............... p. 30
List of Tables

Table 1: Mean IR Reliabilities as a Function of Rating
Instrument and Subject Group ............... p. 31

Table 2: Repeated Measures Analysis of Variance for
Between Subjects Effects (Hypothesis 3) ..... p. 33
Behavioral Specificity and
Reliability in Job Analysis and Job Specification

Job analysis refers to the collection of data describing job-related behaviors and the characteristics of the job environment (Harvey, 1991). Job specification, traditionally considered to fall under the heading of job analysis¹, refers to the process of inferring required traits or abilities necessary for a desired level of job performance. Though job specifications may be based on detailed job analysis data, often incumbents and/or supervisors are simply polled as to what traits they feel are necessary to perform a particular job (cf. Lopez, 1988; Fleishman & Mumford, 1988). The data obtained from job analyses and job specifications are used in a wide variety of personnel functions including compensation, performance appraisal, selection, and training.

This introduction can be divided into two parts: first a broad background segment and a then more focused introductory element. The background segment consists of three parts: (a) an overview of three of the most popular approaches toward job comparison, (b) a bit of background regarding cognitive research, with a focus on schema theory, and (c) a summary of job analysis research to date.

Background: Job Comparison Approaches

The following three approaches will be treated: (a)
the worker-oriented approach embodied in the Position Analysis Questionnaire (PAQ) and Job Element Inventory (JEI), (b) the task-oriented approach developed by the United States Air Force, and (c) the abilities requirements approach made popular by Fleishman and associates and embodied in the Threshold Traits Analysis method.

Although systematic descriptions of jobs have been performed since the 18th century (Primoff & Fine, 1988), generalizable, quantitative comparisons between highly dissimilar jobs have been realized only since the advent and application of factor analysis. The basis for factor analytic studies of jobs was first provided by Lawshe in 1946 (Fleishman, 1967). Since then, a good deal of research has been directed at achieving a single goal: to describe a set of differing jobs along a single dimension or set of dimensions (e.g., Thomas, 1952; Parker & Fleishman, 1961; Fleishman, 1982; McCormick et al., 1957; 1972; Harvey et al., 1988).

In the 1950s McCormick and his colleagues began to gather large amounts of job-related data in hopes of uncovering a single set of dimensions that could be used to describe most, if not all, jobs. Typically, the studies included the following steps: (a) principal components factor analysis performed on a diverse sample of jobs, (b) the derivation from each job of a factor score on each of the resulting factors, (c) the grouping of jobs into levels
on each factor, and (d) the classification of the sample of jobs into patterns in terms of the permutations of all possible factor score levels (cf. McCormick, 1957). As many as 40 or 50 job variables were examined, including aptitudes (e.g., verbal, numerical), physical capacities (e.g., strength, reaching ability), temperaments (e.g., adaptability to repetition, ability to deal with people), interests (e.g., business vs. scientific), work-related behaviors (e.g., reading instruments) and working conditions (e.g., extremes of cold, heat, noise). Literally thousands of jobs were analyzed (mostly by means of incumbents filling out questionnaires) over a period of many years, resulting in the creation of the currently popular instrument, the Position Analysis Questionnaire (PAQ; McCormick et al., 1972). The PAQ is called a worker-oriented instrument, in that jobs are rated and quantified primarily in terms of worker behaviors. PAQ item scores (specific behaviors) are statistically transposed into thirteen underlying work dimension scores, which can then be used for purposes of job comparison. As it is currently administered, PAQ analysis involves the employment of a professional, trained analyst who completes the PAQ questionnaire after interviewing incumbents and supervisors and in some cases directly observing incumbents performing the job under study (McCormick & Jeanneret, 1988). Another worker-oriented instrument, the Job Element Inventory (Harvey et al., 1988),
is designed to be completed by job incumbents themselves.

Also during the 1950s a slightly different approach was developed, refined, and used extensively: task-oriented job analysis (Morsh, 1964). Task-oriented job analysis differs from worker-oriented job analysis in primarily three ways: First, task-oriented items are usually more behaviorally specific than worker-oriented items (e.g., 'monitor radio' vs. 'use communications devices'). Second, item scores are generally not transposed into levels of the more abstract, underlying dimensions of work behavior--item scores themselves are often used for the purposes of job comparison. Third, task-oriented instruments are often developed specifically for a particular group of similar jobs. Worker-oriented instruments, on the other hand, are considered applicable to a wide variety of jobs.

The task-oriented method developed for the United States Air Force (Morsh, 1964) includes the following procedures: Using information derived from job descriptions, training standards, and other available data, a preliminary task inventory is constructed according to a standardized format in an effort to cover the domain of a particular class of jobs (e.g., aircrew member jobs). Subject matter experts (SMEs; e.g., experienced pilots) often act as consultants in the early stages of inventory development to ensure completeness of coverage and accepted use of terminology. The resulting instrument usually
consists of 200 to 300 task statements grouped under major functions which are called duties. The preliminary inventory is submitted to experienced incumbents at several installations who then decide which task statements need to be modified or deleted. New duties and tasks may be added at this stage as well. Finally, the latest revision of the inventory is administered to a large sample of job incumbents. In Morsh's 1964 study it is mentioned that a 3000-case, world-wide survey of incumbents is possible within about 30 days.

A very different perspective is evident in the ability requirements approach developed by Fleishman and associates (Parker & Fleishman, 1961; Fleishman & Mumford, 1988). Rather than focusing on discovering and quantifying job tasks or behaviors, the emphasis is on linking descriptions of job tasks to the general abilities required to perform these tasks. This approach therefore falls under the rubric of "job specification" rather than "job analysis."

Over the past few decades Fleishman and his associates have identified over 40 different abilities relevant to the description of human task performance. These constructs have been derived mainly from the results of a number of factor analytic studies of task inventory data (e.g., Fleishman, 1967; 1978; Fleishman & Quaintance, 1984). Included in Fleishman's taxonomy are three superordinate categories: cognitive abilities (e.g., oral comprehension),
psychomotor abilities (e.g., control precision), and physical abilities (e.g., general hearing).

A typical ability requirements assessment entails the following (Fleishman & Mumford, 1988): A group of subject matter experts, who might include job incumbents, job analysts, or supervisors, are presented with a series of ability rating scales. After a detailed review of the definitions of each ability, the SMEs are then asked to rate the extent to which each ability is required to perform the job under study. Often job descriptions or some other sort of job analysis data are used as a basis for the ratings. Ratings for the job are then averaged across the SMEs to describe and summarize the ability requirements necessary for successful job performance.

Other methods of job comparison include the critical incidents technique (Flanagan, 1954; Bownas & Bernardin, 1988), job element analysis (Primoff, 1957; Primoff & Eyde, 1988), and occupation analysis (Cunningham, 1983, 1988).

**Background: Schema Theory**

A schema can be broadly conceived of as a cognitive structure that is used to receive and organize information (Neisser, 1967). Recently "schema" has also been conceptualized as a knowledge structure consisting of categories, implicit theories, and their process-oriented interrelationships (cf. Rumelhart, 1984; for reviews of schema research, see Alba, 1983; Brewer & Nakamura, 1984).
One of the first schema researchers was Sir Frederick Bartlett who in 1932 published a book entitled *Remembering*, which detailed his studies on patterns of recall. His subjects were first given a collection of stories to read. Bartlett then asked the subjects to recall these stories at various time intervals, ranging up to ten years. Bartlett noted that many of the errors the subjects made were attempts to make the story more coherent and rational—in essence, to follow a understandable pattern or schema. For example, if a particular story contained a passage describing travel by boat, a subject might erroneously remember the activity of fishing as being described in the story (Best, 1986). Bartlett conceived of the schema as "an active organization of past reactions or past experiences" (Bartlett, 1932). He felt that schemas were often used in learning new material and that the subject was more or less unable to separate encoded facts from the previously existing schemas at the time of retrieval (Best, 1986).

Kelly (1951) has also been credited in the development of schema theory. Kelly's "personal construct theory" emphasizes the need to understand how an individual categorizes and acts on social phenomena on the basis of the his or her own implicit personality theory. For example, a perceiver may automatically attribute great leadership qualities to someone who has demonstrated great athletic ability. Such might be the result of an individual's
schematic organization of social perceptions (cf. Mischel, 1986).

About the same time, Piaget developed his theory of child development (Piaget, 1951; cf. Neisser, 1967). According to Piaget, a child enters the world lacking virtually all of the basic cognitive competencies of the adult, and gradually develops these competencies by passing through a series of stages of development (Anderson, 1990). Modern cognitive psychology research has been concerned with identifying the mechanisms that account for these cognitive or perceptual transitions, including how schemas develop and evolve (cf. Anderson, 1990).

Much of the recent work has focused on how schemas organize our interpretation and recall of separate pieces of information. In 1971, for example, Bransford and Franks conducted what is probably the most often cited experiment on schematic memory processes (Hastie, 1981). Their subjects listened to a series of short sentences expressing parts of the meaning of a very long sentence. The long sentence was composed of four interrelated propositions. The sentences to which the subjects were exposed contained from one to three of the propositions related to the single long sentence. Subsequent recognition memory confidence ratings showed that subjects were unable to discriminate between the presented and nonpresented short sentences. The results of the Bransford and Franks experiment represent
evidence for the operation of an interpropositional sentence schema.

Other works have explored various types of categorization schemas (cf. Cantor & Mischel, 1979). For example, the activation or instantiation of a schema has been operationalized in category accessibility research by means of a priming task, in which subjects are stimulated to use a set of trait concepts (Wyer & Gordon, 1984). After priming, the subjects are given information about a target person's behavior that allows substantial flexibility in interpretation. Judgments made about the target are then examined to determine the level of bias toward the particular traits activated by the priming task.

Substantial evidence has surfaced pointing to the existence of prototypical person schemata. Other schema research has explored narrative schemata (e.g., Schank & Abelson, 1977), causal schemata (e.g., Tversky & Kahneman, 1977), and problem-solving schemata (e.g., Taylor, Crocker, & D'Agostino, 1978).

Background: Job Analysis Research

Sechrest (1976) has suggested that the themes of research in his area, personality, change not so much because issues are resolved and phenomena understood, but rather because investigators run out of steam and interest turns to newer, more exciting phenomena. The same statement can be made regarding job analysis research. Aside from the
continuing exploration of the dimensionality of work behavior (described above), there have not been many problems or areas that have been investigated vigorously over a significant amount of time. Several questions have been left unanswered (e.g., which job analysis technique is best?), and those questions that have been answered lie around as "naked facts," that is, there have few attempts to explain them (e.g., the finding that decomposed rating formats are sometimes preferable to holistic rating formats).

One strain of research has received consistent attention over the years and merits a detailed discussion here: the work surrounding Hakel's (cf. Smith & Hakel, 1979) shared-stereotype hypothesis. Other issues to be examined include: (a') ways of assessing a method's validity, (b) variables that may moderate job analysis ratings, and (c) issues relating to rating format (e.g., holistic versus decomposed).

The Shared-Stereotype Hypothesis

The shared-stereotype hypothesis (cf. Smith & Hakel, 1979) maintains, with reference to a particular job under study, that job-naive raters (those with little or no knowledge pertaining to the job) and job-expert raters will hold common stereotypes of the work performed on the job; the two groups will therefore be expected to rate the job similarly. Probably the most serious practical implication
of the hypothesis would be that expert analysts’ ratings could be seen as equivocal, based merely on commonly held cognitive structures rather than on specific job information.

The sources influencing Hakel’s shared stereotype hypothesis are in large measure the early schema theorists mentioned above (e.g., Bartlett, 1932; Kelly, 1951; Piaget, 1951). In addition, leaders in the psychometric or quantitative area (e.g., Cattell 1937; Cronbach, 1955) pointed out the possibility that perceptual responses may depend on stereotypes existing in the mind of the perceiver. In the 1960s such assertions began to acquire a substantial degree of empirical support, particularly in the case of expert judgment (e.g., Chapman & Chapman, 1969). In 1974 Hakel suggested that all rating methods within the personnel area should be reassessed in light of these findings, and in 1979 he and J.R. Smith conducted a study that seemed to support the shared stereotype hypothesis.

The 1979 study involved a PAQ analysis of 25 state government jobs. The raters included groups of job incumbents, supervisors, jobs analysts, and a comparison group of college students. After analysis of the data, Smith and Hakel concluded that there was little difference between analyst sources, including students, in terms of their ability to reliably analyze a job using the PAQ. The conclusion was based largely on the high degree of agreement
shown among all judge categories when summing item frequencies across all 25 jobs.

In 1984 Cornelius et al. pointed out several weaknesses of the study including: (a) differences in item means between the groups were not sufficiently considered, (b) the rater groups were not equivalent in their ability to predict salary levels, (c) the between-group correlations reported were grossly miscalculated (.50s rather than .90s), and (d) the PAQ items rated "Does not apply (DNA)" served to inflate interrater reliabilities and hence obscure differences between the groups. A replication study was also performed, yielding results that failed to support the assertion of equivalence between rater groups.

Other studies conducted in the 1980s have revealed evidence tending to challenge the conclusions drawn by Smith & Hakel. With reference to the problem of DNA items inflating observed correlations, Harvey (1986) demonstrated in a Monte Carlo study that interrater reliabilities in the .50 range can be obtained when raters rule out only 15-20% of the items as DNA and respond randomly to the remainder. In addition, Harvey & Lozada-Larsen (1988), employing the Cronbach (1955) accuracy assessment procedure, found that the amount of job descriptive information available to raters has a significant effect on job analysis accuracy. Raters with more detailed job information were found to consistently make more accurate job ratings than those given
only a job title. In another study DeNisi et al. (1987) found that job-naive raters were unable to yield scores comparable to that of job profiles obtained from PAQ services.

In sum, it can be said that the evidence appears to indicate that job analysis ratings made by job-naive raters are not equivalent to those made by job-expert raters. The shared stereotype hypothesis has been largely disconfirmed, and a deeper understanding of some of the psychometric characteristics of job analysis instruments has been obtained. And while it seems clear that the two groups do not rely heavily on the same job stereotypes, still much remains to be explored in terms of cognitive mechanisms underlying the rating process.

Ways of Assessing Job Analysis Validity

Since validation criteria are difficult or impossible to obtain, it is usually necessary to infer the validity of job analysis data from evidence of reliability as based on results from two or more independent analysts (McCormick, 1979). By far the most popular technique for estimating the reliability of job analysis data has been interrater reliability, usually expressed as the mean of all possible pairwise correlations (Cornelius, 1988). Sometimes researchers will transform the correlations using Fisher's r to Z transformation before the mean is calculated (see Gael, 1988). Occasionally internal consistency measures are
computed (e.g., split-half correlations), treating each rater as an individual item (e.g., Lopez, 1988). Test-retest reliability measures have also been utilized to some extent (e.g., Wilson et al., 1990). Most of the popular job analysis methods have achieved reliabilities ranging from the .70s to the .90s (e.g., Harvey & Hayes, 1986; Wilson & Harvey, 1990; McCormick & Jeanneret, 1988; Morsh, 1964). Unfortunately there is at present no objective way of establishing what can be considered acceptable reliability.

Other approaches toward establishing the validity of job analysis data have had some utility, though have not been employed extensively. One method involves assessing the success to which compensation rates can be predicted. Multiple regression weighting of PAQ data, for example, has resulted in quite powerful predictive capability (cf. Robinson et al., 1974). Another prediction approach involves using job analysis data, particularly derived work dimension scores, to predict how employees in a particular job will score on some other assessment device, such as the General Aptitude Test Battery (McCormick & Jeanneret, 1988). The PAQ has been found to predict cognitive test scores with some accuracy (McCormick & Jeanneret, 1988). This approach is predicated on the assumption that people tend to "gravitate" into those jobs that are commensurate with their own aptitudes. Finally, one technique involves surveying experienced job analysts as to which job analysis method is
best (e.g., Levine et al., 1983; Levine, et al., 1980). The subjective nature of these studies has made interpreting the results difficult.

**Variables That May Moderate Job Analysis Ratings**

Although a number of studies over the years have investigated the effects of potential moderator variables on job analysis ratings, the efforts as a whole can be described as sporadic and desultory. The predominate method of assessing moderator effects has been to analyze data in which variables of interest have been recorded, and then conclude whether such variables had a significant impact on results.

One of the first published studies was conducted by Hazel and associates in 1964. They found distinct differences between the responses made by accountants and those made their supervisors on a task inventory. The two groups agreed only 57 percent of the time on whether or not a particular task was performed. Later studies have also tended to reveal differences between the job perceptions of the worker and the supervisor (e.g., Hauenstein & Foti, 1989).

Other variables explored include degree of incumbent-rater experience (e.g., Landy & Vasey, 1991), level of job performance exhibited by the incumbent-rater (e.g., Conley & Sackett), sex of rater (e.g., Schmitt & Cohen, 1989; Arvey et al., 1977), race of rater (e.g., Schmitt & Cohen, 1989),
and educational level of rater (e.g., Mullins & Kimbrough, 1988; Cornelius & Lyness, 1980). Given the small number of such studies attempted to date, few relationships have been established, except perhaps for the previously mentioned naive rater versus job expert rater distinction.

Two studies conducted by Arvey et al. (1977; 1982) deserve mention on account of the methodological approach employed. In these studies undergraduate students completed job analysis instruments after seeing a color slide presentation and hearing a verbal narrative centered around a hypothetical job. Different treatment groups heard different versions of the narrative, which was manipulated based on certain variables of interest, such as the degree of interest manifested by the hypothetical incumbent. The results of these studies were inclusive and open to question given the artificial nature of the experiments. Nevertheless, an experimental approach such as this has potential for uncovering important aspects of rater behavior and judgmental processes.

**Issues Relating to Rating Format and Procedure**

This area of inquiry is also begging for more comprehensive and systematic programs of research. Certain empirical concerns have been addressed, yet a deeper understanding of the rating process has yet to be aggressively pursued.
As mentioned earlier, the amount of descriptive information provided to the rater has been shown to be related to the accuracy of the ratings (e.g., Harvey & Lozada-Larsen, 1988; Cornelius et al., 1984). There has been some evidence, however, that narrative job descriptions may in some cases be an adequate basis for determinations of job family membership (e.g., Jones et al., 1982; cf. Surette et al., 1990). There certainly must be situations were analyses of a cruder nature are sufficient, though little has been done to address this issue substantively.

Some researchers have argued that holistic measures are sufficient for test validation purposes (e.g., Schmidt et al., 1981), though the available evidence appears to indicate that a "decomposed" judgment strategy, using a mechanical algorithm to determine the overall evaluations, is generally most accurate (e.g., Butler & Harvey, 1988; Cornelius & Lyness, 1980).

Another issue involves the utility of relative-time-spent scales (e.g., 'How often do you perform this task?') versus dichotomous checklists (e.g., 'Do you perform this task?'). With regard to the task inventory approach, recent studies suggest that the two questioning styles possess approximately equal utility (e.g., Wilson & Harvey, 1990; Friedman, 1990).
This Study

Behavior-Trait Linkages

There are potentially important differences between the typical task-oriented or worker-oriented job analysis and the typical trait/ability-oriented job specification (for a full discussion see Harvey, 1991). First, job analysis data, as defined here, is explicitly tied to observable, verifiable job behaviors and job context. For example, a job analysis might involve observational measurement of how often a worker drives a vehicle. Job specification data, on the other hand, involves measurement of hypothetical constructs (e.g., craft skill), entities not explicitly related to any particular job behavior. In this sense job specification is performed on a different level of analysis. This distinction requires that job specifications be based not only on accurate perceptions of work behavior, but also on shared or mutually agreed upon conceptions of specific behavior-trait linkages (see Figure 1).

Consider, for instance, use of a job specification technique known as Threshold Traits Analysis (TTA; Lopez, 1977; Appendix A). This method employs a panel of subject matter experts (SME's), typically either first-line supervisors or experienced incumbents, to serve as raters for a particular job. For each of 33 listed traits, the raters determine, among other things, the level (0-4) of trait necessary for acceptable job performance. From each
A Job Specification Item
(trait/hypothetical construct)

"Craft Skill"

Job Analysis Items:
(frequencies of specific behaviors)

"Operate in Emergency Situations"

"Instruct Others in Some Skill"

"Help People with Problems"

FIGURE 1: An Example of Behavior-Trait Linkages
rater's responses a job/trait profile is derived. With this particular type of approach, a potential problem becomes apparent: Even in the ideal situation in which each rater possesses identical, accurate beliefs regarding the frequency of specific work behaviors the job involves, all raters must share similar conceptions regarding behavior-trait linkages in order for this method to be reliable. All must agree, for example, if the job involves "operating in emergency situations," that this activity implies a particular level of a relevant trait, say "craft skill." This additional source of variability is of some concern, particularly in light of the fact that there is usually substantial disagreement among analysts as to the frequency of specific behaviors a job entails--interrater reliability (to be referred to as "IR reliability"), often used as a measure of validity, ranges from the .70's to .90's (e.g., Harvey & Hayes, 1986; Wilson & Harvey, 1990). Disagreement on behavior-trait linkages has the potential to reduce IR reliability to an unacceptable degree, particularly when a small number of raters are used. With regard to TTA, the number of raters may be as small as five (Lopez, 1988).

**Variance Masking**

Despite this additional variability, however, job specification validation studies have at times registered reliabilities in the high .80's (e.g. Lopez, 1977; Lopez, Kesselman, & Lopez, 1981; Fleishman & Mumford, 1988). As I
will explain, these results can be accounted for by variance masking apparent during the derivation of the trait/ability scores. Let us consider Figure 1, and pretend, for the sake of clarity, that the three specific behaviors listed (e.g., "Operate in Emergency Situations") capture entirely the domain of the trait, "Craft Skill." It is easy to see that the "Craft Skill" score must reflect the extent of each of these behaviors, and thus must represent an aggregation or summing across the behaviors in order to derive a single score. Simplistically we can imagine that the Craft Skill score is an average of the behavior frequency scores of the three specific items. For example if the three behavior frequency scores are 3, 3, and 3 respectively, the Craft Skill score would be 3 as well. If the behavior scores are 2, 3, and 4 respectively (or 1, 3, and 5), the Craft Skill score would still be 3. Such variation in the specific behavior scores is reflected in the IR reliability of behavior specific instruments (i.e., if we only consider the specific behavior frequency ratings), but is not reflected in an assessment of IR reliability of the more condensed or collapsed trait rating instruments (because only the trait ratings are considered). In this way a portion of the variance is masked. The practical implication of this bias, from a psychometric viewpoint, is that when job specifications achieve reliabilities comparable to those found in job analysis, it cannot be concluded that the two
approaches share an equal level of interrater agreement regarding job behaviors. Thus a comparison of job analysis with job specification should not be based solely on measurements of interrater reliability. Rather, this study, drawing upon findings and theory from the schema literature discussed earlier, reflects an attempt to uncover a potential "schema effect."

Schema-based Inference

Quite recently Pryor et al. (1986) examined the influence of schema abstractness upon the processing of social information. They found that the speed with which subjects could classify behaviors as consistent, inconsistent, or irrelevant to a concrete schema (e.g., "brags about accomplishment") was found to be faster than the speed with which they could classify behaviors in relation to an abstract schema (e.g., "is egotistical"). Tentatively they concluded that concrete schematic constructs are more easily utilized in the processing and understanding of social information than more abstract schematic concepts.

While abstract schemas, such as those invoked by job specifications, may lead to more difficult categorization, they may also guide decision-making to a much greater extent than concrete schemas. For example, it has been suggested that abstract, holistic judgments such as those involved in job specification may reflect an overreliance on implicit
theories (Harvey, in press; with regard to performance appraisal, see Feldman, 1981; from the educational research area, see Cadwell, 1986). If such is the case, the popular use of IR reliability as a measure of validity may be misleading. Acceptable levels of reliability may reflect shared schemas rather than actual characteristics of the particular job under study.

This Study

This study investigated (a) the potential schema-based nature of job specification (TTA) ratings, and (b) the relationship between behavioral specificity and IR reliability. Three independent groups of subjects rated the job of college resident assistant (RA). One group consisted of subject matter experts (SME's), in this case RA's who have accumulated over two years of experience at their job. A second group (referred to as the Job-Familiar group) was made up of freshman psychology students who had never performed the job of resident assistant, but had lived under the supervision of an RA for at least one semester. The third group (the Naive group) consisted of freshman and sophomore psychology students who had never worked as an RA or lived under an RA's supervision. All subjects performed a job specification (TTA) and then a worker-oriented job analysis (the Job Element Inventory [JEI]; Cornelius & Hakel, 1978; Appendix B) with regard to the job of RA. In completing these two instruments, the SME subjects were
asked to draw from their substantial job experience, while the Job-Familiar and Naive subjects were instructed to base their ratings on any previous exposure to RA's.

Based on the speculation that (a) job specification involves an extra degree of unreliability due to necessary behavior-trait linkages, and (b) this additional variance is not entirely masked, the following hypothesis was put forward:

1. Within the subject matter expert group, IR reliability will be significantly higher for the JEI than for the TTA.

Hypotheses 2, 3, and 4 made use of the finding that job analysis ratings made by naive raters are not normally as valid as ratings made by subject matter experts (e.g., Cornelius, 1988b; Butler & Harvey, 1988; DeNisi, Cornelius, & Blencoe, 1987). It was expected:

2. With respect to the JEI, IR reliability will be significantly lower for the Naive group than for the subject matter expert group.

The reliability of TTA ratings, on the other hand, may reflect shared convention (schemas) rather than the level of accuracy of job perceptions. It was therefore expected that this instrument's IR reliability would not be as sensitive to a rater's level of job knowledge:

3. Given that the second hypothesis proves correct, the between-group differences in JEI IR reliabilities will be
significantly greater than the between-group differences in TTA IR reliabilities.

Along the same lines of reasoning, it was predicted:

4. The naive group JEI profile (averaged) scores will correlate significantly lower with the SME group JEI profile scores, than will a similar comparison involving TTA profile scores.

Finally, in an effort to further investigate behavioral specificity as it relates to IR reliability, JEI items were divided into two categories based solely on each item's degree of specificity. "Assign people to tasks," for example, was considered more specific than "modify ideas, decisions, or procedures." (Please note that this is a comparison of items within a behaviorally specific job analysis instrument.) Based on the common sense notion that the less specific an item is, the more subject it is to differing interpretations, it was hypothesized:

5. Within the subject matter expert group and with respect to the JEI, IR reliability will be significantly higher for the more-specific items than for the less-specific items.

Method

Subjects

In this study 227 undergraduate students participated. The Job-Familiar group (n=200) and the Naive group (n=10) consisted of freshman and sophomore psychology students.
whose participation was enlisted by means of a sign-up sheet. The psychology student subjects participated in exchange for psychology course credit. At the beginning of the experimental session, the undergraduate subjects indicated on paper whether or not they had lived under the supervision of an RA for at least one semester (designating themselves as either "Job-Familiar" or "Naive"). Though no check was done, it was assumed that the Job Familiar subjects had substantially more exposure to the job of RA than the Naive subjects.

The SME group was made up of 17 resident assistants who participated as part of their duties as RA. A list of potential SME subjects was generated from the pool of active RA's by resident housing administration officials; the criteria used were experience and availability. All SME subjects had performed the job of RA for at least two years. Based on implications contained in the TTA literature (e.g., Lopez, 1977), two years was considered to be the minimum amount of work experience sufficient to merit the designation "subject matter expert." The 17 SME subjects represent persons who agreed to participate from the list of 20 names provided by the resident housing administration. Their support was solicited over the telephone, after they received a letter from their supervisor asking them to participate.
Materials

Two instruments were employed: The Job Element Inventory and the Threshold Traits Analysis Chart. The JEI asks the rater to judge, on a 1 to 5 scale, the relative time spent on each relevant work behavior from a list of 153 items. Deleted from the list were 31 items that could not conceivably apply to a resident assistant (e.g., "Carry firearms"; these items are those that are crossed-through in Appendix B). After the study, the remaining 122 items were divided (for purposes of evaluating Hypothesis 5) into two categories: more-specific (n=37; denoted 'MS' in Appendix B) and less-specific (n=85).

The TTA Chart (Appendix A) asks the rater to determine the level (0-5) of each 33 listed traits necessary for acceptable job performance. A portion of the TTA training materials, presumably those pertaining to detailed explication of the trait labels, were unavailable.

Procedure

The procedure was essentially identical for all groups. The primary difference was that the SME raters were asked to base their ratings on their experience as RA's, while the Job-Familiar and the Naive raters were asked to base their ratings on their prior exposure to the activities of an RA.

A "blind" experimenter administered the two instruments to each group. The psychology student subject groups (Job Familiar, Naive) were tested in trials involving
approximately 30 subjects at a time. In the case of the SME group, trails involved approximately 5 subjects at a time.

After the subjects were greeted, they were asked to complete consent forms. If the trial involved psychology student subjects, they were asked to signify in writing whether or not they had lived for at least one semester under the supervision of an RA. All subjects were then given instruction on how to complete the TTA Chart and had an opportunity to ask questions regarding this technique. Once all the questions had been answered, the subjects were asked to complete TTA Charts with regard to the job of RA. Once all of the TTA Charts had been completed and collected, each subject was given a Job Element Inventory and asked to complete it. Instructions for completing the JEI are contained in the instrument itself, though any pertinent questions were answered as they arose. If it was deemed necessary, the question and answer were reiterated to the entire collection of subjects present. Otherwise, the questions were handled on an individual basis and in an individual, personal manner. As each subject turned in the completed JEI, he or she was given a debriefing statement and thanked for his or her participation.

As outlined above, each subject completed the TTA before completing the JEI. The reason for this ordering (rather than JEI first) lay in the speculation that although TTA scoring may affect subsequent JEI scoring, this effect
is probably not as great as what one might expect when a behavior-specific instrument such as the JEI is completed first, particularly given that the JEI employed may not have covered the entire job domain. It was also believed that if order effects did in fact occur, they were not likely to play a substantial role in the consideration of this study’s primary hypothesis (3), which was assessed by looking at the interaction between the instrument and subject group.

Results

Rather than being converted into z-scores, the IR reliabilities themselves were used in most of the statistical calculations. Figure 2 and Table 1 show the mean IR reliabilities for the six group-by-instrument combinations. Upon viewing Table 1 it becomes apparent that although the two instruments’ respective reliabilities differed in a statistically significant way (in 2 of the 3 groups), there was little practical difference between the TTA IR reliabilities and the JEI IR reliabilities. The greatest difference (.51 vs. .43), occurring within the Job-Familiar Group, cannot be considered very substantial.

However, this difference was found to be statistically significant at the .0001 level—a reflection, most likely, of an inflated number of error degrees of freedom (39798). Unfortunately, a common t-test involving averaged IR reliabilities suffers from a geometric increase in degrees of freedom as the number of raters is increased. A better
Figure 2: Rating Instrument as a function of IR reliability and Subject Group
Table 1

Mean Interrater Reliabilities as a Function of Rating Instrument and Subject Group

A connecting double line (continuous or dashed) reflects a significant difference at the alpha = .05 level.

<table>
<thead>
<tr>
<th></th>
<th>SME Group</th>
<th>Job-Familiar</th>
<th>Naive Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(17 raters</td>
<td>(200 raters</td>
<td>(10 raters</td>
</tr>
<tr>
<td></td>
<td>135 r’s)</td>
<td>19900 r’s)</td>
<td>45 r’s)</td>
</tr>
<tr>
<td>TTA (33 items)</td>
<td>.63 (SD .11)</td>
<td>.51 (SD .16)</td>
<td>.36 (SD .28)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JEI (122 items)</td>
<td>.57 (SD .08)</td>
<td>.43 (SD .19)</td>
<td>.35 (SD .23)</td>
</tr>
</tbody>
</table>
test should be developed employed to handle such a situation.

The primary hypothesis of this study, Hypothesis 3, was tested in an attempt to assess a potential schema effect. "Schema effect," in this case, was operationalized as the two instruments' differential sensitivity to level of rater knowledge. In essence, an interaction was predicted between instrument and rater group. This interaction was evaluated by means of a repeated measures ANOVA, with the 3 rater groups found along the subject dimension, and each instrument serving as one of the two repeated measures. As can be seen in Table 2, the interaction was found to be just barely significant (p = .035). In light of the high error degrees of freedom this test enjoyed (20078), this finding fails to assume practical significance. Thus Hypothesis 3 was not supported—no evidence was uncovered suggesting that the holistic, abstract judgements made during job specifications reflect an overreliance on shared schemas.

Hypothesis 4 was similar in that it also predicted a schema effect. In this case, "schema effect" was operationalized as a difference between the two instruments in terms of the correlation between the Naive and the SME group profile scores. It was speculated that if TTA scores are more dependent on shared knowledge structures rather than specifics of RA job duties, a higher correlation between these subject groups (Naive, SME) should be found in
**Table 2**

Repeated Measures Analysis of Variance for Between Subjects Effects (Hypothesis 3)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Anova SS</th>
<th>Mean Square</th>
<th>F value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP</td>
<td>2</td>
<td>6.07</td>
<td>3.035</td>
<td>82.7</td>
<td>.0001</td>
</tr>
<tr>
<td>ERROR</td>
<td>20078</td>
<td>737</td>
<td>.0037</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Hypotheses for Within Subject Effects

**Source:** INSTR

<table>
<thead>
<tr>
<th>DF</th>
<th>Anova SS</th>
<th>Mean Square</th>
<th>F value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68.5</td>
<td>68.5</td>
<td>3022.9</td>
<td>.0001</td>
</tr>
</tbody>
</table>

**Source:** INSTR*GROUP

<table>
<thead>
<tr>
<th>DF</th>
<th>Anova SS</th>
<th>Mean Square</th>
<th>F value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.152</td>
<td>.076</td>
<td>3.36</td>
<td>.035</td>
</tr>
</tbody>
</table>

**Source:** Error(INSTR)

<table>
<thead>
<tr>
<th>DF</th>
<th>Anova SS</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>20078</td>
<td>455</td>
<td>.023</td>
</tr>
</tbody>
</table>
TTA scores than in JEI scores. At least this was considered a possibility worth investigating. The analysis consisted of first obtaining profile scores by averaging item scores within each of the 2 subject group by 2 instrument combinations. The TTA Naive profile scores were then correlated with the TTA SME profile scores, and the JEI Naive profile scores were correlated with the JEI SME profile scores. These two r's were then converted to z-scores and tested in hopes of revealing a significant difference. The statistical formulas employed are listed in Appendix C. The results were inconclusive at the .05 alpha level (r_{je} = .86; r_{ta} = .91; p = .12). It should be mentioned, however, that DNA (does not apply) responses may have served to inflate the instrument correlations, thereby weakening the power of this test. In any case, as in the Hypothesis 3 evaluation, no conclusive evidence was gathered concerning potential schema effects.

Another equivocal finding was apparent in the analysis of Hypothesis 5, which compared the average IR reliability of JEI more-specific items (n=37) with that of the JEI less-specific items (n=85). A t-test was employed, and only SME ratings were included in the analysis. Again, while the result was statistically significant (p < .0001; error df = 270), the absolute difference (.62 for the more-specific vs. .55 for the less-specific) was not substantial. Given the arbitrary nature of categorizing each item into its
appropriate group (more vs. less specific), it can be confidently stated that considerably more research in this area needs to be undertaken before one may conclude whether or not specificity is positively correlated with IR reliability (within behavior-specific instruments).

Still another marginal and unexpected result occurred with regard to Hypothesis 1. If job specification involves an extra degree of unreliability due to necessary behavior-trait linkages, it was reasoned, the JEI SME group IR reliability should be greater than the TTA SME group IR reliability. Such was not the case, with the TTA achieving .63 and the JEI only .57 (as shown in Table 1). A t-test was employed to assess the difference. Again, the difference was statistically, but not practically, significant (p < .0001; error df = 270). Possible theoretical and psychometric interpretations of this finding are mentioned later in the Discussion section.

Lastly, it can be reported that IR reliability was indeed lower in the JEI Naive group than in the JEI SME group, as was predicted by Hypothesis 2. As can be seen by examining Table 1, however, similar differences occurred in between-group comparisons involving the TTA. For each instrument, a Tukey's HSD comparison was done involving the 3 subject groups (means and SD's are included in Table 1). As illustrated, each subject group was found to be significantly different from the others. Overall, it can be
suggested that the performance of the 2 instruments appears to have differed remarkably little.

Discussion

With one of the five hypotheses clearly supported, the results of this study reveal little evidence to suggest that job specifications cannot attain acceptable reliability or that job specifications are biased in their reliance on shared knowledge structures such as schemas. Contrary to predictions, the worker-oriented job analysis instrument, the JEI, and the job specification instrument, the TTA, achieved similar levels of IR reliability within each of the three subject groups (Hypothesis 1; as shown in Table 1). Also unexpected was the finding that the JEI and the TTA reliabilities varied equally with respect to the level of job familiarity possessed by the respective rater groups (Hypotheses 3 and 4). Both instruments achieved significantly higher reliability in the SME group than in either the Job-Familiar group or the Naive group. This finding suggests that job specifications may be sensitive to a rater's level of job familiarity and that the reliability achieved by job specifications may reflect accurate job behavior perceptions rather than shared, implicit personality theories, categories, or other mechanisms subsumed by the more general term, "schema." As well, though the TTA's IR reliability was modest (.63 for the SME group), it should be pointed out that a substantial portion
of the TTA rater training materials were unavailable at the time of this study. It is possible that the inclusion of these materials would have resulted in reliabilities similar to those reported in earlier studies (e.g., .80's; Lopez, 1977). It is apparent that this study fails to illuminate some of the potential or alleged deficiencies of the job specification approach.

Yet despite the TTA-affirming appearance of these results, alternative interpretations present themselves that merit consideration before an inference can be made regarding the reliability and/or validity of this job specification technique. First, it is likely that there were differences in motivation between the 3 subject groups (SME, Job-Familiar, and Naive). The SME's were described as "go-getters" by their supervisor and participated in this study as part of their job. The other 2 groups participated in order to obtain extra credit in a psychology class. Part of the between-group differences in reliabilities may be explained by the differing degrees of thoughtfulness or seriousness in task completion. It may also be that the Naive subjects felt somewhat confused or put off by being asked to carefully analyze a job they had never been exposed to. Since there was no random assignment to groups, one cannot be sure that the differing TTA reliabilities were solely the result of differences in levels of job familiarity.
Second, to reiterate a point made earlier, the similarity in IR reliabilities between the two instruments may be due to variance masking which occurs during the job specification process. Those completing the TTA must aggregate considerably across behaviors in answering each item. Therefore, the fact that the TTA achieved IR reliabilities comparable to the JEI may be of questionable significance.

Finally, it should be recognized that the JEI performed rather poorly (SME IR reliability of .57) and may not have made for a good comparison. A major problem seems to be that the JEI was developed for the military and validated on military subject samples. 19 percent of the questions had to be deleted (e.g. "Use Mooring Lines") and many of the questions (e.g., "Contact Civilian Specialists") employed a phraseology unique to the military and may have created confusion or misunderstanding among the raters. In hindsight, it may have been better to rewrite the questions in more common language as has been done in a recent study (Harvey et al., 1988).

**Naive Raters versus Subject Matter Experts**

As predicted, and in line with past research (e.g., Cornelius, 1988b), Naive raters were not able to achieve a level of IR reliability comparable to that of the SME raters (Hypothesis 2); it is likely that the Naive ratings were less valid. As mentioned above, however, between-group
differences may be attributable to factors other than job familiarity.

Returning to the issue of schema-usage, it should be recognized that these results do not preclude the possibility of schemas being employed, to at least some degree, by each subject rater. It is possible, for example, that the less job-familiar raters (as in the Naive group) enjoyed a greater diversity of RA-related experiences and hence possessed more individualized, less ingrained, more static conceptions of the job.

The SME subjects were likely to be thinking in quite a different manner. Numerous studies have supported the notion that experts use a more complex and elaborate representation of their knowledge domain (cf. Galambos et al, 1986; Anderson, 1990; Cantor & Kihlstrom, 1989). Expertise does not rest simply on superior perception or memory, but is linked to the detection of familiar and relevant patterns in the stimulus (Gellatly, 1986). Experts also have these elaborate conceptions and representations more readily available or accessible than do novices—chess masters, for example, have described good moves as "just springing off the board" at them (Gellatly, 1986). As a person goes from novice to expert, his or her declarative, factual knowledge of a domain becomes converted into a more efficient procedural representation: A more consistent, automatic reasoning process is often realized. This
fundamental difference between expert and novice cerebration could be why the SME subjects performed differently than the Naive subjects, irrespective of whether the questions were behavior-based (as in the case of the JEI) or trait-based (as in the case of the TTA; Hypothesis 3). The effects of expertise, such as more automatic categorization, may have in fact been enhanced by the more general rather than the more specific questions. This may explain why, within the SME group, TTA reliabilities were higher than JEI reliabilities. It is almost certain, at any rate, that important mental-processing differences existed between the SME raters and the Naive raters, and further job analysis research needs to explore this issue in detail.

Specific Items versus More General Items

As hypothesized, the more-specific JEI items were found to produce significantly higher IR reliability than the less-specific JEI items (Hypothesis 5). However the difference in reliabilities between the item groups was not great (.62 for the more-specific: .55 for the less-specific). Replications of these results involving different instruments and different situations are needed before any strong conclusions can be drawn.

Still a broader question emerges regarding a distinction drawn between the TTA and the JEI. Earlier it was argued that TTA items were hypothetical constructs whereas JEI items were specific behaviors. Upon reflection
it appears harder to make a great distinction between the more general JEI items (e.g., "Modify Ideas, Decisions, or Procedures") and the TTA items, which include both an ability/trait and a behavioral description of that ability/trait (Appendix A). As Binning & Barrett (1989) argue, viewed pragmatically, a construct can be seen as merely a hypothesis about which behaviors will reliably covary. Seen in this light, this distinction between job specification and job analysis may not be entirely clear-cut. In fact in this particular case, the distinction appears to be a rather fuzzy one.

**Future Research**

Job analysis researchers should continue to explore social cognition and related literatures in an effort to better understand the rating process. By importing and integrating theory and methodology from these areas we may be better able to assess the strategies used to capture the job domain. We may, for example, come upon better ways of assessing an instrument or technique’s validity. If we rely on measures of reliability (such as IR reliability) to gauge an instrument’s validity, we will remain open to the charge that consensus reflects collective bias rather than shared accuracy (see Chan, 1982; Chapman & Chapman, 1969). This study represents an attempt to unearth or lay bare such potential bias (characterized, in this study, as schema-effects). It is hoped that future efforts will also address
this problem, perhaps utilizing a more sophisticated or powerful methodology.

One methodological aspect that allows considerable room for improvement concerns distinguishing raters along a specific subject variable, such as level of job familiarity. More extensive subject classification procedures in the future might involve completion of a "job exposure checklist." In the study just completed, the psychology student subjects were merely polled as to whether they had spent a semester living under the supervision of an RA. Future classifications might include questions such as "Does your knowledge of job duties come from first-hand experience (and, to what extent), friends who have had this job, "word of mouth," or, alternatively, no definable origin?" A more extensive survey of subjects may more clearly operationalize the concept, "naive rater."

It is also hoped that this study will draw attention to the need for an inexpensive, JEI-type job analysis instrument that is well-suited for civilian use. Optimally, such an instrument would be designed for those on a modest reading level (e.g., high school reading level), and perform adequately in validation studies using incumbent raters. Development should include multiple validation criteria (e.g., IR reliability, ability to predict compensation rates, ability to cluster jobs within job families), and adequate validity should be established over a range of
occupational venues. Too many companies today operate without the aid of job analysis information, and such an instrument may offer a low-cost way to enhance the performance of personnel management systems.

Finally, I would urge more exploration of ways to link specific job behaviors to specific job skill or ability constructs, as was done in a recent study by Hughes & Prien (1989). There has been a call for more emphasis on criterion development (Binning & Barrett, 1989), and establishment of such linkages will help by reducing the inferential leap personnel specialists must make in selecting predictor constructs and in determining performance criteria. In this way organizations will make the best use of behavior-specific job analysis information.
References


Footnotes

1I've accepted the argument for a necessary distinction between the terms "job analysis" and "job specification." For a full discussion see Harvey, 1991.
The TTA Trait Questionnaire

PLEASE Take Plenty of Time.

Each answer is important.

Write your answers in the blanks provided
I am:

A. a person who has lived under the supervision of an RA.

B. a person who has never lived under the supervision of an RA.

If you put "A," For how long? _____________________

PLEASE WAIT FOR INSTRUCTIONS BEFORE CONTINUING
The TTA Trait Questionnaire

PLEASE Take Plenty of Time.

Each answer is important.

Write your answers in the blanks provided
ABILITIES

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Strength</td>
<td>Lift, pull or push physical objects</td>
</tr>
<tr>
<td>2. Stamina</td>
<td>Expending physical energy for long periods</td>
</tr>
<tr>
<td>3. Agility</td>
<td>React quickly, has dexterity, coordination</td>
</tr>
<tr>
<td>4. Vision</td>
<td>See details and color of objects</td>
</tr>
<tr>
<td>5. Hearing</td>
<td>Recognize sound, tone and pitch</td>
</tr>
<tr>
<td>6. Perception</td>
<td>Observe and differentiate details</td>
</tr>
<tr>
<td>7. Concentration</td>
<td>Attend to details amid distractions</td>
</tr>
<tr>
<td>8. Memory</td>
<td>Retain and recall ideas</td>
</tr>
<tr>
<td>9. Comprehension</td>
<td>Understand spoken and written ideas</td>
</tr>
<tr>
<td>10. Problem-solving</td>
<td>Reason and analyze abstract information</td>
</tr>
<tr>
<td>11. Creativity</td>
<td>Produce new ideas and products</td>
</tr>
<tr>
<td>12. Numerical Computation</td>
<td>Solve arithmetic and numerical problems</td>
</tr>
<tr>
<td>13. Oral Expression</td>
<td>Speak clearly and effectively</td>
</tr>
<tr>
<td>14. Written Expression</td>
<td>Write clearly and effectively</td>
</tr>
<tr>
<td>15. Planning</td>
<td>Project a course of action</td>
</tr>
<tr>
<td>16. Decision-making</td>
<td>Choose a course of action</td>
</tr>
<tr>
<td>17. Craft Knowledge</td>
<td>Apply specialized information</td>
</tr>
<tr>
<td>18. Craft Skill</td>
<td>Perform a complex set of activities</td>
</tr>
</tbody>
</table>

CONTINUED ON NEXT PAGE
<table>
<thead>
<tr>
<th>Attitude Traits</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>19. Adaptability--Change</td>
<td>Adjust to interruptions and changes</td>
</tr>
<tr>
<td>20. Adaptability--Repetition</td>
<td>Adjust to repetitive activities</td>
</tr>
<tr>
<td>21. Adaptability--Pressure</td>
<td>Adjust to critical and demanding work</td>
</tr>
<tr>
<td>22. Adaptability--Isolation</td>
<td>Work alone or with little personal contact</td>
</tr>
<tr>
<td>23. Adaptability--Discomfort</td>
<td>Work in hot, cold, noisy work places</td>
</tr>
<tr>
<td>24. Adaptability--Hazards</td>
<td>Work in dangerous situations</td>
</tr>
<tr>
<td>25. Control--Dependability</td>
<td>Work with minimum of supervision</td>
</tr>
<tr>
<td>26. Control--Perseverance</td>
<td>Stick to a task until completed</td>
</tr>
<tr>
<td>27. Control--Initiative</td>
<td>Act on own, take charge when needed</td>
</tr>
<tr>
<td>28. Control--Integrity</td>
<td>Observe regular ethical and moral codes</td>
</tr>
<tr>
<td>29. Control--Aspirations</td>
<td>Limit desire for promotion</td>
</tr>
<tr>
<td>30. Personal Appearance</td>
<td>Meet appropriate standards of dress</td>
</tr>
<tr>
<td>31. Tolerance</td>
<td>Deal with people in tense situations</td>
</tr>
<tr>
<td>32. Influence</td>
<td>Get people to cooperate</td>
</tr>
<tr>
<td>33. Cooperation</td>
<td>Work as a member of the team</td>
</tr>
</tbody>
</table>
APPENDIX B
Instructions for BEHAVIOR method

This next questionnaire is very different. You will be asked to rate how often a specific BEHAVIOR is performed. Forget about traits!

The job to be rated is still Resident Advisor (RA).

A 5 point rating scale will be used, but it is very different. In using the rating scale, you will select a number between 1 and 5 that best describes the RELATIVE TIME SPENT on each BEHAVIOR. RELATIVE TIME SPENT means "amount of time spent engaged in that behavior compared to typical behavior for that job."

For example, one would probably rate a statistician as a "5" on the item "USE QUANTITATIVE MATERIALS." This is because a statistician spends a lot of time using quantitative materials (numbers, charts, graphs, etc.).

IMPORTANT:
If an item does not apply to this job being rated (Resident Advisor), then please leave it blank. For example, "USE QUANTITATIVE MATERIALS" would be left blank for the job of ditch digger.

If you have any questions and at any time, just raise your hand. And take your time--going too fast hurts your brain and hurts our results. Thanks!
The Job Behavior Questionnaire

PLEASE take plenty of time on each question

If you get finished before the others, take a second to look over your responses

EACH ANSWER IS IMPORTANT!
RATING SCALE: Relative Time Spent
1 2 3 4 5
Very Below Average Above Very
Little Average Average Much

U.S. COAST GUARD ENLISTED PERSONNEL JOB ELEMENT INVENTORY

If the job element is not appropriate, LEAVE IT BLANK.

In doing your job, you...

1. USE WRITTEN MATERIALS (tech manuals, notices) ___ (15)
2. USE QUANTITATIVE MATERIALS (graphs, tables of numbers) ___ (16)
3. USE PICTURES OR PICTURE-LIKE MATERIALS (blueprints, maps) ___ (17)
4. USE PATTERN DEVICES (templates, stencils, radio codes) ___ (18)
5. USE VISUAL DISPLAYS (gauges, radarscope) ___ (19)
6. USE PHYSICAL MEASUREMENT DEVICES (rulers, pressure gauges) ___ (20)
7. USE FEATURES OF NATURE (cloud formations, stars, ocean disturbance) ___ (21)
8. USE MAN-MADE FEATURES (bridges, dams, docks) ___ (22)
9. USE VERBAL COMMUNICATIONS ___ (23)
10. USE SOUNDS (engine sounds, sonar) ___ (24)
11. USE TOUCH ___ (25)
12. USE ODOR (applies to any odor you need to smell to do your job) ___ (26)
13. USE TASTE (food preparation) ___ (27)
14. PERFORM TASKS THAT REQUIRE YOU TO SEE EXTREME DETAIL OF OBJECTS (reading small print, setting ignition points) ___ (28)
15. PERFORM TASKS THAT REQUIRE YOU TO SEE MODERATE DETAILS OF OBJECTS (hammering nails, reading gauges) ___ (29)
16. PERFORM TASKS THAT REQUIRE TREATMENT OF SICK OR INJURED ___ (30)
RATING SCALE: Relative Time Spent

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very Below Average</td>
<td>Average</td>
<td>Above Average</td>
<td>Very Much</td>
<td></td>
</tr>
</tbody>
</table>

If the job element is not appropriate, LEAVE IT BLANK.

In doing your job, you...

17. WORK OUTDOORS

18. WORK IN AN ENCLOED AREA THAT IS HOT

19. WORK IN AN ENCLOED AREA THAT IS COLD

20. WORK IN POLLUTED AIR (dust, toxic fumes)

21. ARE SUBJECT TO VIBRATION

22. WORK UNDER IMPROPER LIGHTING CONDITIONS (too dark, too glaring)

23. WORK WHERE YOU EASILY BECOME DIRTY

24. WORK IN A CRAMPED OR UNCOMFORTABLE SPACE

25. WORK IN A QUIET AREA

26. WORK IN AN AREA OF MODERATE NOISE (office with typewriters)

27. WORK IN AN AREA OF LOUD NOISE

28. ARE RESPONSIBLE FOR THE SAFETY OF THE GENERAL PUBLIC

29. ARE RESPONSIBLE FOR THE SAFETY OF MEMBERS OF THE COAST GUARD

30. JUDGE DISTANCES

31. TELL THE DIFFERENCE IN COLORS

32. NOTICE DIFFERENT PATTERNS OF SOUND ( Morse code, engines not running right)

33. NOTICE DIFFERENCES OR CHANGES IN SOUND THROUGH LOUDNESS, PITCH OR TONE QUALITY

34. SENSE BODY POSITION AND BALANCE (walking on I beams, walking on deck)

CC

(31)

(32)

(33)

(34)

(35)

(36)

(37)

(38)

(39)

(40)

(41)

(42)

(43)

(44)

(45)

(46)

(47)

(48)
RATING SCALE: Relative Time Spent

1 2 3 4 5
Very Below Average Above Very Little Average Average Much

If the job element is not appropriate, LEAVE IT BLANK.

In doing your job you...

35. JUDGE SPEED OF MOVING OBJECTS

36. JUDGE SPEED OF SOME PROCESS (cooking time, developing pictures)

37. INSPECT PRODUCTS, OBJECTS, MATERIALS OR EQUIPMENT

38. JUDGE SIZE OR WEIGHT OF OBJECTS WITHOUT DIRECT MEASUREMENT

39. GATHER OR ARRANGE INFORMATION INTO A MEANINGFUL ORDER

40. CODE AND DECODE ( Morse code, computer language)

41. MAINTAIN LOGS

42. SUBTRACT, MULTIPLY, AND DIVIDE NUMBERS

43. WORK WITH PERCENTAGES, FRACTIONS, AND DECIMALS

44. USE ALGEBRAIC, GEOMETRIC, TRIGONOMETRIC, AND STATISTICAL METHODS

45. USE TOOLS THAT PERFORM PRECISE OPERATIONS

46. USE TOOLS WITH LONG HANDLES (brooms, shovels)

47. USE TOOLS OR DEVICES FOR THE PURPOSE OF HANDLING THINGS (tongs, ladles)

48. USE HAND-HELD POWERED DEVICES THAT PERFORM VERY PRECISE OR ACCURATE OPERATIONS (soldering irons, welding equipment)

49. USE HAND-HELD POWERED DEVICES LIKE POWER SAWS AND DRILLS

50. USE DEVICES THAT YOU DRAW OR WRITE WITH

51. USE DEVICES THAT APPLY SOMETHING (brushes, paint rollers)

52. USE MOoring OR Towing LINES

53. USE STATIONARY MACHINES OR EQUIPMENT THAT YOU CONTROL

67
RATING SCALE: Relative Time Spent

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very Little</td>
<td>Below Average</td>
<td>Average</td>
<td>Above Average</td>
<td>Much</td>
</tr>
</tbody>
</table>

If the job element is not appropriate, LEAVE IT BLANK.

In doing your job, you...

54. USE DEVICES THAT HAVE FIXED OR VARIABLE SETTINGS (TV selector switch, room thermostat) ____(68)___

55. USE KEYBOARD DEVICES (adding machines, typewriters) ____(69)___

56. USE SMALL BOATS ____(70)___

57. DRIVE CARS OR TRUCKS ____(71)___

58. USE WHEELBARROWS AND LAWN MOWERS ____(72)___

59. USE REMOTE CONTROLLED EQUIPMENT ____(73)___

60. SET UP OR ADJUST MACHINES OR EQUIPMENT ____(74)___

61. USE HANDS DIRECTLY TO FORM OR CHANGE MATERIALS ____(75)___

62. TAKE EQUIPMENT APART OR PUT IT BACK TOGETHER ____(76)___

63. ARRANGE OR PACK OBJECTS OR MATERIALS ____(77)___

64. PERFORM TASKS THAT REQUIRE HIGHLY SKILLED BODY COORDINATION ____(1)___

65. PRESENT INFORMATION TO PUBLIC GROUPS ____(2)___

66. PERFORM TO ENTERTAIN (band) ____(3)___

67. ATTENDING TO OTHERS' NEEDS (waiving on tables, cutting hair) ____(4)___

68. CONTACT FLAG OFFICERS, AND OTHER HIGH OFFICIALS AS PART OF MY JOB ____(5)___

69. CONTACT OFFICERS IN GRADE OF LCC, COR, CAPT AS PART OF MY JOB ____(6)___

68
If the job element is not appropriate, LEAVE IT BLANK.

In doing your job, you...

72. CONTACT CIVILIAN PROFESSIONALS AS PART OF MY JOB (doctors, lawyers, professors, engineers) (9)

73. CONTACT CIVILIAN SPECIALISTS (draftsmen, designers, photographers, law enforcers, meteorologists) (10)

74. CONTACT MERCHANT MARINE AND OTHER INDUSTRY PERSONNEL (11)

75. CONTACT OTHER GOVERNMENT PERSONNEL (Army, Navy, Air Force) (12)

76. CONTACT CIVILIAN SALESMEN AND SUPPLIERS (13)

77. CONTACT FOREIGN NATIONALS (14)

78. CONTACT PUBLIC (boating safety, environmental protection, law enforcement) (15)

79. CONTACT CIVILIAN STUDENTS (16)

80. CONTACT SPECIAL INTEREST GROUPS (property owners, boating clubs, local governments) (17)

81. SUPERVISE NON-COAST GUARD PERSONNEL (18)

82. OPERATE IN EMERGENCY SITUATIONS (19)

83. DEAL WITH PEOPLE IN DIFFICULT SITUATIONS (EEO and drug problems, law enforcers) (20)

84. TAKE RISKS WHILE SERVING OTHERS (SAR teams) (21)

85. PERFORM IN DANGEROUS SITUATIONS (22)

86. PERFORM THE SAME PHYSICAL TASK OVER AND OVER (23)

87. PERFORM THE SAME MENTAL TASK OVER AND OVER (24)
RATING SCALE: Relative Time Spent

| Very Little | Below Average | Average | Above Average | Very Much |

If the job element is not appropriate, LEAVE IT BLANK.

In doing your job, you...

88. WORK ON A SCHEDULE THAT ALLOWS YOU SOME FREEDOM AS LONG AS YOU FINISH YOUR JOB

89. FOLLOW CERTAIN SET PROCEDURES ON YOUR JOB (like following a check-out list to inspect equipment)

90. PERFORM UNDER TIME PRESSURES

91. CONTINUALLY WATCH FOR EVENTS THAT HAPPEN RARELY IN YOUR JOB BUT ARE IMPORTANT OR CRITICAL

92. CONTINUALLY WATCH FOR FREQUENT CHANGES IN YOUR JOB SITUATION (rescue traffic, constantly watching gauges and dials that change often)

93. WORK UNDER DISTRACTIONS

94. MOVE LIGHT OBJECTS ON OCCASION

95. MAKE EFFORTS ABOUT EQUAL TO LIFTING 25 TO 50 POUNDS

96. MAKE EFFORTS ABOUT EQUAL TO LIFTING 50 TO 100 POUNDS

97. USE FINGER MOVEMENTS (drawing instruments, keyboard devices)

98. PERFORM TASKS THAT REQUIRE A STEADY HAND AND ARM

99. COORDINATE HAND AND/OR FOOT MOVEMENT WITH WHAT YOU SEE (driving a car, steering a boat)

100. COORDINATE YOUR HAND MOVEMENTS WITH WHAT YOU HEAR

101. ADVISE PEOPLE IN RESOLVING THEIR PROBLEMS

102. PERSUADE OTHERS TOWARD SOME ACTION OR OPINION

103. INSTRUCT OTHERS, FORMALLY OR INFORMALLY, IN SOME SKILL OR KNOWLEDGE
RATING SCALE: Relative Time Spent

<table>
<thead>
<tr>
<th></th>
<th>Very</th>
<th>Below Average</th>
<th>Average</th>
<th>Above Average</th>
<th>Very Much</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Little</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the job element is not appropriate, LEAVE IT BLANK.

In doing your job, you....

104. ANALYZE PROBLEMS

105. ANSWER QUESTIONS FROM OTHERS

106. ANTICIPATE THE NEED FOR MATERIALS TO ACCOMPLISH WORK

107. ANTICIPATE THE NEED FOR MANPOWER TO ACCOMPLISH WORK

108. APPROVE REQUESTS AND/OR PROPOSALS FROM OTHERS

109. SETTLE DISPUTES AMONG OTHERS

110. SCHEDULE MEETINGS AMONG PEOPLE

111. ASSESS THE QUALITY OF WORK OF OTHERS

112. ASSIGN PRIORITIES TO TASKS

113. ASSIGN PEOPLE TO TASKS

114. GIVE FORMAL BRIEFINGS TO OTHERS

115. CLARIFY GOALS AND TASKS FOR OTHERS

116. COMPILE DATA FOR DECISIONS

117. DEMONSTRATE TECHNIQUES AND PROCEDURES

118. MODIFY EQUIPMENT

119. MODIFY IDEAS, DECISIONS, OR PROCEDURES

120. MONITOR EQUIPMENT

121. MOVE EQUIPMENT AND SUPPLIES

122. PREDICT FUTURE EVENTS

123. PREPARE PLANS AND SCHEDULES
RATING SCALE: Relative Time Spent

If the job element is not appropriate, LEAVE IT BLANK.

In doing your job, you...

124. PRESIDE OVER MEETINGS ____ (61)
125. RECOMMEND PROCEDURES AND COURSES OF ACTION ____ (62)
126. RESOLVE CONFLICTING FINDINGS ____ (63)
127. REPAIR EQUIPMENT ____ (64)
128. ACQUISITION EQUIPMENT ____ (65)
129. SELECT APPROPRIATE EQUIPMENT AND MATERIALS ____ (66)
130. SUPERVISE OTHERS ____ (67)
131. DISCUSS ISSUES AND PROBLEMS WITH OTHERS ____ (68)
132. DISSEMINATE INFORMATION TO OTHERS ____ (69)
133. DISTRIBUTE EQUIPMENT AND SUPPLIES ____ (70)
134. DRAFT WRITTEN MATERIALS ____ (71)
135. DRAW UP PLANS OF ACTION ____ (72)
136. ENCOURAGE THE EFFORTS OF OTHERS ____ (73)
137. ENFORCE DIRECTIVES ____ (74)
138. ESTIMATE TIME, COST AND OTHER NEEDS FOR PROJECTS ____ (75)
139. FORMULATE POLICIES ____ (76)
140. IDENTIFY CAUSES OF EQUIPMENT PROBLEMS ____ (77)
### Rating Scale: Relative Time Spent

<table>
<thead>
<tr>
<th>Very Little</th>
<th>Very Average</th>
<th>Above Average</th>
<th>Very Above</th>
<th>Much</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

If the job element is not appropriate, leave it blank.

In doing your job, you...

| 141. Identify causes of personnel problems | 1 |
| 142. Are accountable for decisions and actions of others | 2 |
| 143. Supervise the operation of equipment | 3 |
| 144. Test equipment | 4 |
| 145. Provide first aid | 5 |
| 146. Dispense medication | 6 |
| 147. Stand watch | 7 |
| 148. Consider ideas and suggestions of subordinates | 8 |
| 149. Adjust to new situations | 9 |
| 150. Keep supervisor informed | 10 |
| 151. Serve as fireman | 11 |
| 152. Carry firearms | 12 |
| 153. Maintain records | 13 |
Appendix C

Formula for converting r's into z-scores:
\[ z = 0.5 \ln \left( \frac{(1 + r)}{(1 - r)} \right) \]

test of significance (Hypothesis 4):
\[ z_{obs} = (z_{ta} - z_{je}) / \sqrt{k} \]
where \( k = \frac{1}{(n_1 - 3)} + \frac{1}{(n_2 - 3)} \)

RR: \( z_{obs} > 1.645 \)
Marc Cowgill

Addresses

Home

609 Clay St. Apt. 5
Blacksburg, VA 24060
ph. 703-953-0476

Permanent

5997 Queenston St.
Springfield, VA 22152
ph. 703-569-0416

Personal Information

Born: Oxnard, California
Date of Birth: July 25, 1962

Educational Background

1987 B.A. in Psychology, University of Virginia,
Charlottesville, VA

Work Experience

Present Graduate Assistant for the Virginia Polytechnic
and State University Psychology Department.

1987-1988 Cook at Pizza Hut on Emmet Street,
Charlottesville, VA.

1986-1987 Animal technician, University of Virginia,
Charlottesville, VA

1984-1986 Supervisor at Public Interest Communications,
Inc., Falls Church, VA. Duties included
supervision of a telephone fund-raising
operation.

Community Service

Special reading tutor for foreign elementary school
students in Blacksburg, VA, 1990

Foreign Language

Knowledge of German.
Special Interests

Job Analysis
Personnel Psychology
History of Psychology
Computer Programming
English Vocabulary
Progressive Music

Awards and Honors

Bestowed a 4 year Presidential Fellowship by the Virginia Polytechnic and State University, 1987

Special Competencies

Teaching

Areas in Psychology
Statistics

Computer Programming

Pascal and Basic

References*

Dr. R. J. Harvey
VPI and SU
Blacksburg, VA 24060
703-231-7030

Dr. Roseanne Poti
VPI and SU
Blacksburg, VA 24060
703-231-5814

Dr. Neil Hauenstein
VPI and SU
Blacksburg, VA 24060
703-231-5716

Dr. Sigrid Gustafson
VPI and SU
Blacksburg, VA 24060
703-231-3133

*Additional references available on request.

August 1991