PLACE IN RETURN BOX to remove this checkout from your record. TO AVOID FINES return on or before date due. MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

6/01 c:/CIRC/DateDue.p65-p.15

A QUANTITATIVE REVIEW OF PREDICTORS OF JOB TASK AND CITIZENSHIP PERFORMANCE

By

Brian Hahn Kim

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Department of Industrial/Organizational Psychology

ABSTRACT

A QUANTITATIVE REVIEW OF PREDICTORS OF JOB TASK AND CITIZENSHIP PERFORMANCE

By

Brian Hahn Kim

A large body of research on job performance has examined citizenship performance behaviors in contrast with job task behaviors. However, findings in the literature have not always provided consistent and overwhelming support for contextual, or citizenship, performance theories, particularly with regard to its hypothesized determinants. Meta-analytic results of this study partially support stipulations of Motowidlo, Borman, and Schmit's (1997) revised theory of contextual/citizenship job performance. Personality dimensions tended to predict citizenship behaviors better than task behaviors. However, cognitive ability remained the single best construct-level predictor across the performance dimensions, in a variety of settings. Biodata also predicted both task and citizenship performance very well. The implications of using such a two-dimension framework are discussed.

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
INTRODUCTION	1
Job Performance	5
Task Performance	6
Citizenship / Contextual Performance	8
Going beyond formal job tasks	11
Jobs confounding citizenship and task performance	13
Challenging behaviors	15
Recipients of citizenship: people, tasks, or organizations	17
Relative importance of task and citizenship	18
Integrated frameworks of citizenship performance	20
Empirical Evidence for the Task-Citizenship Distinction	22
Job Performance Antecedents	26
Cognitive Ability	30
Personality	34
Structured Interviews	37
Biodata	39
Summary of Research Hypotheses	41
Conclusion	46
METHOD	48
Literature Search	48
Criteria for Study Inclusion	48
Data Coding Procedure	49
Interrater Agreement	51
Meta-analytic Procedure	59
Corrections for Artifactual Variance	63
Measurement unreliability	63
Range variation	65
Multivariate Meta-analysis Procedure	69
RESULTS	74
Database Description	74
Outlier Analyses	77

Overview of Meta-analytic Results	79
Hypothesis 1: Moderation by Job Type	89
Hypothesis 2: Moderation by Citizenship Dimension	93
Hypotheses 3 and 4: Differential Prediction Patterns for Performance	
Dimensions	98
Hypotheses 5 and 6: Moderation by Job Complexity	99
Hypotheses 7 through 10: Specific Predictior – Criterion Relationships	101
Hypothesis 12: Prediction of Biodata Linked to Constructs	104
Supplemental Analyses	106
DISCUSSION	111
Review of Research Goals	111
Summary of Findings	112
Future Directions	118
Limitations	122
Overall Conclusion	124
APPENDIX A: Pilot Coding Sheet	126
APPENDIX B: Coding Sheet	138
APPENDIX C: Code Book	140
	140
APPENDIX D: Interrater Coding Agreement Results	144
APPENDIX E: SAS / IML program for multivariate computations	145
APPENDIX F: Full List of Studies in Database	151
APPENDIX G: Scree Plots for Outlier Analyses	154
APPENDIX H: Job Complexity Codes	164
	101
APPENDIX I: Biodata Studies For Which Raters Assigned Construct Codes	165
APPENDIX J: Results for Restricted Samples	166
	100
REFERENCES	167

LIST OF TABLES

Table 1. Study Variable Labels and Definitions	52
Table 2. Reliability Information For Scales	56
Table 3. Database Descriptives	76
Table 4. Descriptive Statistics of Reliability Estimates	78
Table 5. Meta-analytic Correlation Matrix for Job Performance and Performance Predictors	82
Table 6. Meta-analytic Results For Pairs of Study Variables	86
Table 7. Estimated Population Correlation Matrix Based on Multivariate Meta-analysis	92
Table 8. Tests of the Moderating Effect of Job Type	95
Table 9. Percentage of Correlations From Each Personality Dimension	95
Table 10. Simple Comparisons of Job Dedication and Interpersonal Facilitation.	97
Table 11. Tests of the Moderating Effect of Job Complexity	102
Table 12. Pairwise Meta-analytic Estimates for the Multivariate Sample	110
Table 13. Summary of Conclusions for Hypotheses.	113

LIST OF FIGURES

Figure 1. Model of Hypothesized Relationships Between Predictor Variables and Job Performance Dimensions	44
Figure 2. Model of Hypothesized Relationships Between Personality Dimensions and Citizenship Dimensions	45
Figure 3. Multilevel Structure of Meta-analytic Database	62
Figure 4. Path Diagram with Multivariate Estimates of Study Variables	91

INTRODUCTION

After nearly a century of concerted research on the topic, I/O psychologists continue to disagree on what exactly constitutes job performance (Campbell 1990; Coleman & Borman, 2000; Rotundo & Sackett, 2002; Van Dyne, Cummings, & Parks, 1995) despite the notion that "individual performance on a 'task,' virtually any task that the culture views as having value, is one of the most important dependent variables in psychology, basic or applied" (Campbell, McCloy, Oppler, & Sager, 1993, p. 35). At the most general level, job performance is the set of behaviors executed by an employee in the context of work that contribute to the overall effectiveness of an organization. Understanding which employees perform this set of behaviors well and how they do so is a central imperative for many areas of industrial-organizational research, most notably in areas of personnel selection and job training.

Despite a general recognition that people can and do perform many different actions at work, examinations of distinct aspects of job performance have been largely ignored in research (Austin & Villanova, 1992). Job performance is most often measured globally with a supervisor or peer rating (Bernardin & Beatty, 1984; Cascio, 1995; Cleveland, Murphy, & Williams, 1989; Scullen, Mount, & Goff, 2000). The use of broad performance measures is practical and efficient for making simple evaluations based on the rank order of individuals, as in selection. Furthermore, composite scores of performance based on more specific measures tend to distribute people according to a compensatory model of performance in which few workers excel on every dimension.

Despite these advantages, however, broad measurements of overall performance result in a loss of information about specific causal relationships and typically leave a

considerable portion of variance to be explained in the criterion (e.g., Schmidt, 2002; Schmidt et al., 1985; Viswesvaran & Ones, 2002; Schmidt & Hunter, 1998). Measures of overall performance can also create conceptual ambiguity when based on lower-order constructs that are quite different and caused by different factors. That is, the antecedents (e.g., cognitive ability) having the greatest influence on a performance composite may differ from those having the greatest influence on a specific aspect of performance. There remain some large gaps in the research literature about the nature of specific relationships between various predictors and the possible facets of job performance (Hough & Oswald, 2000). This problem has been evident for some time in "the criterion problem" that results from the reliance on a single, criterion deficient outcome (Austin & Villanova, 1992; Nathan & Alexander, 1988). To the extent that different aspects of job performance are uniquely influenced by various determinants, theory requires critical examinations of the full range of performance antecedents and how they cause each performance behavior or process.

Fortunately, a few researchers have already attempted to lay the conceptual foundation needed to define and study job performance at a more detailed level. Around the late 1980s and early 90s, a number of job performance taxonomies were put forth by various parties (e.g., Brief & Motowidlo, 1986; Campbell 1990; McCloy, Campbell, & Cudeck, 1994; Organ, 1988; Schmidt & Hunter 1992). Campbell et al. (1993; Campbell, Gasser, & Oswald, 1996) developed a broad taxonomy of performance behaviors that was intended to account for all jobs in the *Dictionary of Occupational Titles*. The taxonomy consists of eight components including aspects of task proficiency, demonstrating effort, maintaining discipline, facilitating peer performance, supervising,

and managing. The Campbell model can be applied to many situations and has undoubtedly sparked interesting research questions that would be ignored by focusing on overall performance.

Although the general growth of large taxonomic systems like that proposed by Campbell and colleagues has improved our ability to create theories about work processes, a fair amount of attention has been focused on a particular set of behaviors that were not traditionally thought of as job performance. Although not always well defined, the general class of behaviors related to creating and maintaining a positive work environment for the purpose of enhancing an individual's or group's capability of producing organizational output seems important for making workers efficient and satisfied, and achieving organizational success. Such behaviors have, in fact, been included in a number of job performance models (e.g., Borman & Motowidlo, 1993; Campbell et al., 1993; Organ, 1988; Van Dyne et al., 1995) and typically are distinguished from other behaviors related more directly to the production of organizational output.

While the 1990s witnessed the development of a number of labels for these behaviors gain popularity, such as contextual, organizational citizenship, prosocial, personal initiative, and extra-role performance, a recent confluence of research has begun to support the various investigations of environment-supporting, or contextual, behaviors as one broad dimension of job performance that is distinguishable from task-related behaviors performed explicitly to deliver organizational output. The consideration of a general distinction between "task" and "contextual" behaviors allows for a more focused study of performance at a level of detail just one step removed from the use of overall

performance measures, and is believed to benefit past performance models by increasing parsimony and generalization across jobs.

Borman and Motowidlo (1997, Coleman & Borman, 2000) offered their initial (1993) model of contextual performance as an overarching framework that subsumes many facets of performance generally related to enhancing the work environment, or context. Campbell and colleagues (1996) found their model to be compatible with the Borman and Motowidlo (1993) framework by directly linking a portion of their dimensions to contextual performance. Others (e.g., Organ, 1997; Podsakoff, MacKenzie, Paine, & Bachrach, 2000; Rotundo & Sackett, 2002) have linked the concept of contextual performance to a set of almost parallel concepts (Organ, 1988) broadly termed "organizational citizenship behaviors."

This study reviews the conceptual framework of contextual, or (as it is now commonly labeled) citizenship, performance as posited by previous theorists. As more than a decade has passed since the introduction of the task-citizenship performance taxonomy, it is appropriate to evaluate how this perspective has influenced research and whether the distinction has elucidated our understanding of performance behaviors. This review also helps to identify conceptual ambiguities and potential levers for closing the theoretical gaps with research. Based on the literature review, a series of hypotheses concerning the differential patterns of relationships between the task and citizenship performance dimensions and various performance predictors are then tested in a quantitative review of relevant published research. This study extends findings from previous work by providing meta-analytic estimates of cumulated data, by examining commonly used, practical measures of performance predictors as well as measures of

theoretical constructs, and by simultaneously considering the relationships of various predictors of task and citizenship performance across a wide range of jobs.

Job Performance

Despite its obvious importance, research on the concept of performance had been largely absent in the literature before the late 1980s (Campbell et al., 1996; Ford, Kraiger, & Schechtman, 1986; Motowidlo, Borman, & Schmit, 1997). Initially, performance was assumed to be a general, uniform construct and a sufficient outcome against which other phenomena could be validated. However, the general definition that has become prevalent today considers performance to be those *behaviors, under the control of the individual, that contribute to the goals of an organization employing the individual* (Campbell et al., 1996; Murphy & Cleveland, 1995). Though not necessarily restricted to observable behaviors, performance is made up of actions and not intentions or consequences of those actions, as stated in the definition (Campbell et al., 1993; Murphy, 1989). Furthermore, performance is not equivalent to effectiveness. Job performance is simply action in the context of work; it can be executed well or poorly. Effectiveness and other outcomes have a valence attached to them, a valence beyond the individual's control.

Beyond the definition, it is generally agreed that job performance consists of too many distinct behaviors to be considered a single theoretical construct. The idea that everything a person does at work (that contributes to organizational effectiveness) is the same thing, job performance, with the same antecedents and consequences is hardly useful. An analogous, equally impotent perspective would be to label most human behaviors as simply life performance. Therefore, there has always been a need to

differentiate performance behaviors based on their relationships to other constructs in a nomological network for improving the theoretical meaning and practical usefulness of job performance concepts.

Task Performance

Most would agree that any definition of job performance should at least include those tasks that provide essential functions for transforming an organization's raw input to output (Borman & Motowidlo, 1993; Campbell et al., 1993; Rotundo & Sackett, 2002), without which the organization would not survive. The foundation of modern I/O psychology was, in fact, spurred on by task-based work such as Taylor's initial studies of scientific management (Taylor, 1911, 1912; Locke, 1982). With a lack of theory about human resource potential, a readily available (and disposable) workforce, and burgeoning interest in assembly line-type work, there was little point in studying non-task behaviors during that era. Starting with those early studies, a plethora of research has supported the validity and use of formal job tasks in understanding the nature of work (Austin & Villanova, 1992).

Borman and Motowidlo (1993) conceptualized "task performance" as *the activities that execute or indirectly service core technical functions that transform environmental resources into organizational products*. Task behaviors cover a wide range of behaviors performed by workers at all levels and can range from assembling car parts to taking customer service calls to planning inventory shipments. Other theories of job performance also seem to endorse this basic premise (Rotundo & Sackett, 2002). A major focus on task measures as performance criteria makes intuitive sense – they represent the "core functions." For instance, National Research Council's Committee on

the Performance of Military Personnel views work samples as the only true measure of performance because they demonstrate actual behaviors required on the job (see Campbell et al., 1996). However, the centrality of task performance in research becomes questionable when considering the many complex processes that occur in actual job settings.

Early researchers recognized the importance of "other" individual characteristics that were important for accomplishing job tasks, including one's effort level, sense of loyalty, and willingness to be helpful and cooperate (Barnard, 1938; Katz, 1964). Jobs today typically consist of a number of varied tasks that are strung together and performed over time, often in different or changing environments (Cascio, 1995). Organizations are also increasingly using team-based structures where interdependent goals can only be achieved through multiple people working in concert with each other (Kozlowski, Gully, Nason, & Smith, 1999; Lawler, Mohrman, & Ledford, 1995). To successfully adapt and coordinate their tasks with others, workers may find it necessary to focus on non-task related behaviors such as generally getting along with coworkers or deferring personal responsibilities to backup or "cover for" someone who is unable to perform at a particular time (Blickensderfer, Cannon-Bowers, & Salas, 1997; Dickinson & McIntyre, 1997).

It is also likely that non-task behaviors become increasingly beneficial after one becomes familiar with core tasks. As job tasks become easier and require less attention through practice and/or automatization, workers can focus on other aspects of work such as taking the initiative to perform additional work, finding ways to improve core functions, or helping coworkers with issues unrelated to the job. These non-task behaviors, in turn, may increase organizational effectiveness. Even the famed Hawthorne

project produced results that are in line with this notion. Worker productivity increased not only because workers were given more attention but because they were able to create a supportive social environment. Task performance may also play a smaller role in upper-level jobs where managers are concerned with the general state of the organization and production, organizational politics, managing others who perform core functions, defending the organization, and more, rather than performing core tasks themselves. Thus, task performance is very important but does not appear to capture the entire domain of behaviors that lead to organizational effectiveness.

Citizenship / Contextual Performance

Organizational research has increasingly focused on behaviors that improve the general social and psychological context in which job tasks are performed. These contextual, or citizenship, behaviors are believed (e.g., Borman & Motowidlo, 1993; Campbell et al. 1993) to be important dimensions of a worker's overall contribution to an organization and are believed to have a set of determinants that is unique from that for task performance.

Unfortunately, there was very little consensus about what specific activities comprise this alternative dimension of job performance initially, and many ambiguities still remain. Nonetheless, people have suggested that behaviors generally supporting the work environment, or context, are important and worth studying. Among the specific behaviors investigated in recent years are following rules, volunteering to do extra or unrelated work, showing extra effort and perseverance, or defending organizational objectives from external criticism. Almost as many performance concepts as there are behaviors have been offered by researchers to classify types of contextual activities,

including but not limited to contextual performance, citizenship performance, organizational citizenship behavior (OCB), prosocial behavior, personal initiative, loyalty, interpersonal facilitation, and whistle-blowing. Some efforts to integrate concepts and theories have been made but the model of task and contextual performance by Borman and Motowidlo (1993) is, perhaps, the broadest and most flexible, enough so to serve as an overarching framework for the various types of performance listed above (see Borman & Motowidlo, 1993, 1997 for explanations of the various concepts; Coleman & Borman, 2000), subsuming or being synonymous with many terms.

Contextual performance (later renamed citizenship) was originally defined as *the activities that support main task functions by shaping the organizational, social, and psychological context in which they are carried out* (Borman & Motowidlo, 1993). Restated, contextual performance is the set of activities that are under the individual's discretion and contribute to organizational effectiveness but are not task performance. These behaviors differ from task performance in four basic ways (Borman & Motowidlo, 1993). First, they do not support the technical core itself as much as its environment. Consequently, being proficient is less important than demonstrating initiative beyond a base level of requirements or expectations. Second, they are common to all jobs, unlike core tasks that vary by job and organizational goal. Third, their variance is largely determined by volition and predisposition rather than by knowledge, skills, and abilities (KSA's) leading to proficiency. Fourth, they are "not likely" to be required or explicitly rewarded by a role though sometimes formally recognized in certain jobs. In addition, this facet of performance appears to be more affective (Hattrup, O'Connell, & Wingate,

1998; Motowidlo & Van Scotter, 1994; Penner, Midili, & Kegelmeyer, 1997) or attitudinal in nature (Organ, 1997; Penner et al., 1997) than task performance.

Borman and Motowidlo (1997) revised their contextual performance conception by tying it to existing frameworks of OCB, soldier effectiveness, sportsmanship, whistleblowing, courtesy, civic virtue, and employee reliability to develop a five-category taxonomy of behaviors: persisting with effort and enthusiasm, volunteering beyond one's job tasks, helping and sportsmanship, following rules and civic virtue, and endorsing organizational objectives. They admit that their distinction between contextual and task performance remains blurred but held to the following three assertions: 1) important task activities differ by job while important contextual behaviors generalize across jobs (e.g., being amicable is helpful for a salespeople and machinists), 2) tasks are "more likely" to be role prescribed, and 3) individual differences in task performance are determined more by cognitive ability while differences in contextual performance are determined more by personality. Motowidlo and others (1997) later specified a more specific model where the links between personality and cognitive ability with task and contextual performance were mediated by relevant skills, habits, and knowledge, also reiterating their third assertion.

While the use of the Borman and Motowidlo model as an overarching framework seems plausible and useful, a number of conceptual debates over the nature of different aspects of citizenship performance must be addressed before this middle range concept can be applied to integrate theory and test expectations with real world phenomena. Four debates of particular importance have generated much discussion among researchers and are addressed here. First, the idea of extra-role behaviors is distinguished from

citizenship performance. Second, the importance of citizenship in managerial jobs is addressed. The third debate compares and contrasts helping behaviors with "challenging" behaviors. The fourth separates interpersonal and task aspects of citizenship. The section is concluded with a brief discussion about the interaction between task and citizenship performance before an attempt is made to form a unified concept of citizenship performance that is comparable to the meso-level variable of task performance.

Going beyond formal job tasks. Overlapping with citizenship performance, OCBs are defined as the behaviors across time and persons that jointly promote organizational effectiveness but are not formally required or directly rewarded (Organ, 1988); they are extra-role and discretionary (Borman & Motowidlo, 1997). Organ (1997) later reexamined his definition and dropped the requirement that OCBs are extra-role, realizing that classification schemes should not label a particular behavior differently depending on the setting in which it is viewed. A salesperson may be required to smile when a customer enters the building but is performing no differently than a custodian who smiles, despite not being required to do so. This led Organ to conclude that OCB is "synonymous" with contextual performance but to promote the continued use of his term, OCB, because he "find[s] that both academic and practitioner types readily and intuitively grasp what it is all about" (p. 91) and because the term contextual performance "simply strikes [him] as cold, gray, and bloodless" (p. 91). Despite this rather blithe justification for labeling, Borman and others (e.g., Borman, Penner, Allen & Motowidlo, 2001 and Coleman & Borman, 2000) appear to have converted, using the term "citizenship performance." However, it is duly noted that citizenship performance is not

strictly equivalent with contextual performance because Borman and Motowidlo (1997) include whistle blowing and similar behaviors as part of contextual performance while Organ does not.

Similarly, Van Dyne and colleagues (1995) defined extra-role behaviors (ERBs) as behaviors that are intended to or do benefit the organization, are discretionary, and go beyond existing role expectations. This definition has obvious relevance to the extent that it overlaps with contextual performance but is distinct in two ways. First, it relies on subjective perceptions about what is "required" by a role, either explicitly or implicitly. These perceptions can vary for superiors who rate performance and for workers who decide whether or not to perform beyond their formal or perceived role. Second, the term ERB includes behavioral intentions and implies the notion of altruism, requiring that a person's intention is to help the organization or another person and not merely oneself. While the latter stipulation can be useful in understanding and predicting certain actions (Hogan, Rybicki, Motowidlo, & Borman, 1998), it is irrelevant from the perspective of actually measuring performance. People who act in a certain way are said to be performing, regardless of their intentions. Conceptually, the inclusion of altruistic behavioral intentions is thus incompatible with most accepted definitions of performance that are defined in terms of behaviors (e.g., Campbell et al., 1993) – requiring them as the base unit of analysis. Practically, a focus on ERBs will undoubtedly lead to frequent measurement errors as the same behavior can be labeled differently depending on how observers infer someone's intentions and make causal attributions about actions (Schnake, 1991). For example, the act of complimenting could be seen as supportive and cooperative for one person and as ingratiating and sly for another. Therefore, this study

will be limited to understanding explicit performance behaviors but does recognize that the concept of intentions and perceiving responsibilities beyond formal, role requirements may be important in other work.

Jobs confounding citizenship and task performance. Some believe that task performance in managerial and service jobs is confounded with citizenship behaviors because these workers spend a considerable portion of their time nurturing the social environment of coworkers and less time dealing with core production (Borman & Motowidlo, 1993) than other types of workers. The assumption is that the primary function of managers is to provide social support and endorse/protect the organization, and that corresponding acts comprise their main job tasks, many of which are likely to be formally required. The theoretical implication is that managers' "task" performance is simultaneously citizenship performance. The practical implication is that organizational attempts to increase or enhance managerial performance will have the same effect on citizenship and on task performance (Conway, 1999).

Conway (1999) suggested that interpersonal activities related to guiding and developing subordinates who perform tasks were, in essence, the managerial version of task performance, in addition to technical-administrative duties that were more directly related to core production. As when discussing the extra-role distinction, this type of reasoning is questionable according to the behavioral definition of performance since the same behavior should not be labeled differently just because it is performed by a manager rather than by a lower-level worker (Organ, 1997; Rotundo & Sackett, 2002). Instead, activities tied closely to the administration of core production processes, direction about how to plan and organize production, and backing up subordinates to enable production

should be considered task performance. Activities that are further removed from core functions, such as showing loyalty to a group, helping subordinates with personal issues, defending the organization, and demonstrating loyalty, would fall in the domain of citizenship performance. Admittedly, Conway (p. 5) does state that the key distinction rests in whether or not managerial behaviors are "more explicitly oriented toward goal achievement." Although there may be some "gray area" in distinguishing such behaviors, reframing the debate in this way allows us to ask more meaningful and testable questions.

Perhaps managerial jobs require more citizenship performance relative to task performance (cf. Ilgen & Hollenbeck, 1991). Conway (1999) found support that both job dedication and interpersonal dimensions of citizenship contributed uniquely to overall managerial performance, beyond the contributions of task performance. The arguments above could similarly be applied to sales employees who must maintain a supportive environment to achieve sales (MacKenzie, Podsakoff, & Fetter, 1991). Vinchur, Schippman, Switzer, and Roth (1998) found that conscientiousness was a good predictor of sales criteria while cognitive ability only predicted ratings criteria well. If, as many hypothesize, conscientiousness enables workers to take initiative, put forth extra effort, and be dedicated to their job, then it should be related to performance in sales jobs where the usefulness of citizenship is more salient. Ultimately, the belief is that managers or sales agents are expected, or required, and rewarded to perform citizenship behaviors well, thereby increasing the correlation between citizenship predictors and job performance criteria.

Hypothesis 1 (H1): Citizenship performance will show higher positive correlations with noncognitive predictors in managerial and sales jobs than other jobs.

However, the effect predicted by Hypothesis 1 might not occur for two specific reasons. First, there may be little observed variance in citizenship performance by managers if they were selected on their aptitude or willingness to perform citizenship behaviors or if citizenship behaviors are formally required. Second, studies with poorly defined criteria that label citizenship behaviors as task performance will most likely attenuate any correlations between citizenship performance and noncognitive predictors. The second effect is controlled for in this study by using a set of rules for categorizing performance criteria into task and citizenship dimensions based on definitions of performance derived from the literature (described in the Methods section).

Challenging behaviors. This debate concerns the difference between behaviors that promote an organization and behaviors that challenge it (Van Dyne et al., 1995; Organ, 1997). Challenging refers to behaviors like whistle blowing, principled organizational dissent, and general voice. While conflicting definitions have emerged in the literature, whistle blowing generally refers to discretionary behaviors that disclose an illegal, immoral, or illegitimate act with the intention of ultimately improving the organization (Van Dyne et al., 1995). Persons can benefit personally, but only in addition to their contribution to the organization, though they are often penalized for their acts. For example, the whistleblowers in three recent scandals (i.e., Enron, WorldCom, and FBI) initially attempted to rectify problems internally and privately (without public recognition) before deciding that a more drastic measure was necessary to invoke

changes (Lacayo & Ripley, 2002). Principled organizational dissent is opposition to practices that are not illegal but are still objectionable on the basis of "conscientious principles." Voice behaviors promote change rather than prohibit current practices; they may include persuading others, counteracting groupthink, or providing constructive criticism. Challenging then refers to a broader group of behaviors "criticizing the inefficiency of the status quo" for the benefit of the organization (Van Dyne et al., 1995, p. 252).

Puffer (1987) stated, "noncompliant behaviors are distinct types of nontask behavior that have a common achievement-motivation base but are influenced by different perceived situational contingencies" (p. 619). Compared to citizenship, challenging behaviors appear to have a "different character altogether," sometimes incurring immediate costs before eventually benefiting the organization (Organ, 1997). At the same time, challenging appears to affect organizations through the psychological environment more than through job tasks (Borman & Motowidlo, 1997), and appears to be determined by personality and motivation more so than by cognitive ability (LePine & Van Dyne, 2001). For these reasons, it is included here under the broad category of citizenship, though future investigations can and should assess the extent to which challenging is different from other dimensions of citizenship. Finally, it is noted that challenging behaviors are arguably distinct from sheer negative or retaliatory acts such as sabotage or counterproductive behaviors that are not related to achieving organizational goals (Kelloway, Loughlin, Barling, & Nault, 2002; Miles, Borman, Spector, & Fox, 2002; Puffer, 1987; Rotundo & Sackett, 2002).

Recipients of citizenship: people, tasks, or organizations. Citizenship behaviors are typically performed for certain targets or recipients. The recipient may be the organization or an individual coworker. Some argue for separating OCB-I (behaviors directed at other individuals) and OCB-O (behaviors directed at the organization). OCB-Is may be determined more by personality, trust, and emotional expression whereas OCB-Os may be determined by conscious, cognitive decisions to reciprocate in a social exchange (Lee & Allen, 2002; Settoon & Mossholder, 2002). LePine, Erez and Johnson (2002) found little support for this distinction. Due to a dearth of empirical support for this distinction in the current literature, these behaviors will be treated similarly in this paper.

Van Scotter and Motowidlo (1996) split the concept of citizenship in a similar fashion into two dimensions: interpersonal facilitation and job dedication. Interpersonal facilitation refers to social acts of helping and cooperating with others while job dedication refers to self-disciplined motivated acts like working hard, taking initiative, and following rules. In their study, results suggested that job dedication is not clearly distinct from task performance. The constructs were moderately correlated (r = .48) and had similar patterns of relationships with experience, ability, job knowledge, and personality (average r = .15 with conscientiousness). This led those authors to believe that motivational elements related to task performance account for part of the citizenship domain. In contrast, results supported interpersonal facilitation as being unique from both job dedication and task performance (r = .36 and .35, respectively). Johnson (2001) similarly concluded that a (motivational) measure of job-task conscientiousness was related to aspects of both task and contextual performance.

Organ and Ryan (1995) conducted separate meta-analyses for altruism and generalized compliance when estimating the correlations between predictors and OCBs based on theoretical grounds. As the patterns of relationships were similar across the two analyses and no estimate of the correlation between the two dimensions of citizenship was provided, the usefulness of this distinction still needs to be investigated. Hurtz and Donovan (2000) found similar patterns of relationships when the Big Five personality dimensions were correlated with job dedication and with interpersonal facilitation. The only exception was for agreeableness, which was slightly more related to interpersonal facilitation ($r_c = .20$) than job dedication ($r_c = .10$).

Though this partitioning of citizenship is tentative based on the existing empirical evidence, there is enough speculation that interpersonal behaviors directed at other people may be different from behaviors directed at the organization or work to warrant further investigation. The primary analyses of this study are concerned with the difference between task and citizenship but a secondary hypothesis concerning this moderator of correlations involving citizenship is that:

Hypothesis 2 (H2): Effect sizes for measures related to citizenship performance will be moderated by the degree to which interpersonal facilitation and job dedication aspects are measured.

Relative importance of task and citizenship. A corollary to the task-citizenship distinction is that all organizations require task performance by definition, or else there is no "work" to be done. Some minimum level of task output (basic competency or "satisficing") is inevitably required for an organization to exist. Conversely, organizations do not necessarily require individuals to exhibit citizenship behaviors (e.g., fully automated systems or small organizations with little interaction between individuals). This reasoning might explain the finding by Rotundo and Sackett (2002) that citizenship is weighted more for effective task performers. So it may be the case that a minimum level of task performance must be demonstrated before any worth is attributed to a worker.

However, citizenship behaviors may allow organizations to reach maximal levels of effectiveness or to ensure continuous development or survival. The clearest examples of citizenship that are likely to help individuals function together above typical levels of task performance are conscientious behaviors and following rules. By using the social and psychological environment to support core functions in this way, workers (particularly managers) presumably enable the organization as a whole to function in an integrated fashion that is better than the sum of individual performances.

Citizenship concerns may also take precedence over core functioning for pragmatic reasons, becoming the source of individual differences when job applicants all perform at similar levels of task performance (e.g., simple jobs that anyone can do well). Alternatively, organizations requiring strictly routinized task performance with little room for discretion may find that certain aspects of citizenship contribute little to overall effectiveness or are detrimental because they distract workers from their task performance (Hunt, 2002). Similarly, task performance may also have greater practical utility in extremely complex jobs with varying assignments where the majority of an individual's attention and effort must be devoted to perform a particular task well (e.g., aeronautical engineers and physics professors). In conclusion, both concepts may have

great practical and theoretical importance but an absence of task performance logically precludes the need for citizenship performance.

Integrated frameworks of citizenship performance. There have been a few attempts at integrating different models of performance to form a unified theory of citizenship. Van Dyne and colleagues (1995) proposed a complex nomological network for classifying four extra-role constructs: OCB, prosocial behaviors, whistle blowing, and organizational dissent. They were able to substantively clarify constructs by reducing conceptual overlap in previous definitions. Among their recommendations, they suggest concentrating on citizenship behaviors as a broader and more consistent term. (The reader is referred to the original article for specific conclusions – many of which are regarded as irrelevant here due to their focus on ERBs rather than citizenship/contextual aspects.)

Coleman and Borman (2000) derived an overall model of citizenship performance by using factor analysis, multidimensional scaling (MDS), and cluster analysis on a similarity correlation matrix composed of citizenship dimensions that were sorted by I/O psychologists. The factor analysis produced four factors that accounted for fifty-nine percent of the variance and were interpreted as: 1) helping and cooperating with others, 2) endorsing, supporting, and defending the organization, 3) following organizational rules, and 4) persisting with enthusiasm and extra effort to complete own task. The MDS analysis resulted in five groups of behaviors: 1) interpersonal altruism, 2) interpersonal conscientiousness, 3) organization allegiance/loyalty, 4) organizational compliance, and 5) job/task conscientiousness. Complementing this, the cluster analysis supported three groups of behaviors: 1) interpersonal citizenship performance, 2) organizational

citizenship performance, and 3) job/task conscientiousness. The authors concluded that the analyses together support three broad categories of citizenship performance depending on who or what benefits from a behavior: the whole organization directly (e.g., endorsing, supporting, following rules), other workers (e.g., helping, cooperating), or the job/task (e.g., conscientiousness, extra effort). As addressed earlier in the section about recipients of citizenship actions, there is a lack of support for categorizing behaviors explicitly directed at the organization, precluding a meta-analysis. So, this study only tests the distinction between interpersonal behaviors and other (job dedication and taskdirected) behaviors in H2, per the recommendations of Coleman and Borman (2000) and Van Scotter and Motowidlo (1996).

Rotundo and Sackett (2002) reviewed the array of related concepts in the literature and concluded that definitions of citizenship performance continue to overlap and rely on "muddied" features of behavior (e.g., extra role, not explicitly rewarded, or formal part of the job). They recommended defining performance behaviors independent from the context in which they are performed or from their consequences. They then treated citizenship performance as a single, broad concept.

LePine and others (2002) similarly supported a broader level of analysis, stating that specific dimensions of citizenship can be treated as equivalent indicators of a common latent construct, "a general tendency to be cooperative and helpful in organizational settings" (p. 61). In their study, potential sub-dimensions of citizenship showed high intercorrelations, similar relationships with predictors, and little incremental variance when compared to a measure of general citizenship, and nonsignificant moderators.

Based on these qualitative and quantitative reviews, this study relies on a general theory of citizenship performance that reverts to Borman and Motowidlo's original definition of a general class of behaviors that support the social and psychological environment. Overall job performance then appears to consist mostly of task and citizenship dimensions. Facets of citizenship behaviors may be conceptually distinct but can be treated similarly because they are likely to be determined by the same constructs and contribute to organizational effectiveness in a similar manner. However, the past literature does suggest that citizenship behaviors might be composed of two distinguishable facets at an intermediate level of detail: interpersonal facilitation and self-disciplinary acts.

The theory also implies that citizenship and task performance will be determined primarily by different individual characteristics. Task performance should be more strongly related to cognitive ability while citizenship performance should be more strongly related to personality and motivation. Also, though the two dimensions of job performance are distinct and probably weakly related, task performance may be viewed as having more weight for the survival of an organization. Finally, citizenship performance is separated from extra-role behaviors that may or may not be characterized as aspects of the work environment rather than of behaviors that are tied to core functions.

Empirical Evidence for the Task-Citizenship Distinction

Since the early 1990s, empirical support for the distinction between task and citizenship behaviors has accumulated. If these concepts are to be useful in theory, they must show distinct patterns of relationships with other variables in a nomological

network (Cronbach & Meehl, 1955). If they are to be useful in practice, they must also be related weakly to each other or else they will be functionally redundant. Overall, the literature seems to support the task-citizenship distinction using both of these standards.

Motowidlo and Van Scotter (1994) examined supervisor ratings of task, contextual, and overall performance in relation to experience ability, training, and personality in a sample of Air Force mechanics. Both task and contextual performance predicted incremental variance in overall performance over each other (within an estimated range of reliabilities, .4 to .8). Task ratings explained between 17% and 44% of the variance in overall performance above contextual ratings; contextual ratings, between 12% and 34% above task ratings. Each criterion also produced a different pattern of correlations with individual characteristics, where personality correlated more strongly with contextual ratings than task ratings. Unfortunately, conclusions in this study were questionable because the data failed to show an expected large correlation between task performance and cognitive ability. The authors also limited the generalizability of their conclusions based on the idea that military jobs might involve discretionary behavior infrequently as compared with civilian jobs.

With confirmatory factor analyses of fifteen multitrait-multirater matrices, Conway (1996) showed that a task/contextual model of performance fit better than a unidimensional one, particularly for nonmanagerial performance ratings. Correlations within a domain tended to be higher than between domains; mean correlations were .70 (SD = .11) for task-task, .70 (SD = .13) for contextual-contextual, and .55 (SD = .15) for task-contextual ratings across raters. Conway also examined whether contextual performance subdimensions were differentially related to task performance. He

concluded that three subdimensions were distinct, finding that cooperating had a lower correlation with task performance (.51) than did following rules (.72) or persisting with extra effort (.59). Finally, there appeared to be no differences in reliability between task and contextual measures. Together, the findings based on these ratings supported a distinction between the two performance dimensions (Borman et al., 2001).

Hattrup and colleagues (1998) studied the relationships of cognitive ability and conscientiousness with sales performance, absenteeism, tardiness, and OCBs in a sample of sales representatives over a 6 month period. Cognitive ability was significantly correlated only with sales performance (.31) while conscientiousness was related to absenteeism (-.24) and OCBs (.23) but not tardiness. The same pattern was found when examining the incremental validity of either predictor. Their results supported the general notion that task and contextual performance are different aspects of overall job performance.

For eight job families in a telecommunications firm, Johnson (2001) showed that interpersonal citizenship, organizational citizenship, job-task conscientiousness, and handling work stress all explained incremental variance in overall performance above dimensions of task performance. Other research has produced similar results when predicting other broad outcomes like systemic rewards and promotability (Allen & Rush, 1998; Van Scotter, Motowidlo, & Cross, 2000). Johnson also found that task criteria exhibited a pattern of relationships with cognitive ability and personality that was different from citizenship criteria, except for job-task conscientiousness which appeared relevant to both performance dimensions.

There is also some evidence showing that ratings of overall performance in the literature have included aspects of citizenship performance. Lance and Bennett (2000) evaluated a structural equation model of supervisory performance ratings made by Air Force personnel. Their model fit the data, where performance ratings were mediated by aspects of task and contextual performance. Rotundo and Sackett (2002) used a policycapturing approach to assess the weights given to task, citizenship, and counterproductive behaviors in managerial ratings of overall performance. Raters sorted descriptions of hypothetical workers in five job types with respect to the three performance subdimensions. Managers appeared to rely primarily on three weighting strategies that varied across job types. One group of managers weighted task performance most highly. Another group weighted task and citizenship nearly equally. The third group gave counterproductive ratings the most weight, followed by task and then citizenship behaviors. There was also a significant interaction between task and citizenship ratings in predicting overall performance, suggesting that managers value citizenship more in workers who accomplish their job tasks.

Conclusions about counterproductive behaviors in this study are problematic for two reasons. First, the authors scaled the worker profiles to reflect each type of performance equally, creating a nonrealistic worker in light of the low base rate of counterproductive behaviors that occur in real work environments. Second, the operational definition of counterproductive behavior was not necessarily distinguishable from poor citizenship performance (e.g., low levels of compliance, a facet of contextual performance, were defined as counterproductive). Because such behaviors fall on a continuum of levels, extremely poor compliance is likely to hurt the organization whereas

much compliance can help it. Thus, they defined the same type of behavior differently depending on its level.

In conclusion, past research supports the task-citizenship distinction. Despite this, we still lack a consistent estimate of the magnitude of the relationship between task and citizenship/contextual performance which determines the usefulness of the distinction in many applications like the measurement of individuals in selection or training. Estimates that exist in the literature cover a wide range. Hattrup and collaborators (1997) estimated the correlation between task and contextual performance at .18, using figures from Day and Silverman (1989), Motowidlo and Van Scotter (1994), and McHenry, Hough, Toquam, Hanson, and Ashworth (1990). In contrast, Murphy and Shiarella (1997) set the correlation between task performance and OCBs at .00 based on a brief summary of conflicting estimates. Hattrup and other colleagues (1998) found task (sales) performance to be nonsignificantly correlated with absenteeism, tardiness, and OCB (-.18, -.16, .19, respectively). Higher estimates include those by Allen and Rush (1998) at .66, Johnson (2001) at .54, and Beaty, Cleveland, and Murphy (2001) at .75. As a result, one primary goal of this study is to estimate the relationship between task and citizenship performance with quantitative summary methods.

Job Performance Antecedents

Regardless of whether one is referring to overall job performance or specific dimensions, complex phenomena such as performance are likely to be multiply determined. Campbell and colleagues (1993, 1996) proposed that individual differences on job performance components are completely determined by declarative knowledge, procedural knowledge and skill, and motivation. That is, specific job performance

behaviors may be caused by various influences including factors intrinsic to a person or to environmental factors, but variance between individuals who perform the same task will be determined by the three components.

Motowidlo and others (1997) applied this general theory to a more detailed model in which "habits, skill, and knowledge" mediate the influence of cognitive ability and personality on task and contextual performance dimensions, with personality being more strongly related to contextual performance and cognitive ability being more related to task performance. These conclusions were based on a brief qualitative review of empirical findings related to the prediction of task and contextual performance with personality and general cognitive ability.

Yet, many different explanations of this hypothesized pattern can be found in various works, most of which seem to be based on intuition. People who are motivated to contribute to organizational effectiveness can do so through either task or contextual means, or through both. Because task performance can be complex, cognitive ability explains many differences between individual performances. Motivational and dispositional factors related to being conscientious are also believed to affect task performance to some degree, particularly when extra effort or care is needed to ensure successful production. In contrast, theories of citizenship have implied that social skills, personality, and motivation are strong determinants of differences between workers. These antecedents enable or motivate certain people to be of greater assistance or to make extra effort in all aspects of their work, under the assumption that citizenship behaviors are common across jobs and are not necessarily easier to perform for people of higher cognitive ability. In addition, some have proposed the idea that individuals (Hogan et al., 1998; Penner et al., 1997) choose to increase citizenship behaviors when task productivity cannot be increased; there is more discretion for performing citizenship behaviors than for task behaviors, discretion that is determined by personality. This may occur either when a worker lacks the ability to improve task performance or when work processes are very structured and not amenable to improvements.

It is also likely that the motivation to be generally helpful and cooperative is rooted in one's personality, affective disposition, or current mood state (Beaty et al., 2001; Day & Silverman, 1989; Gellatly & Irving, 2001; McHenry et al., 1990; Motowidlo & Van Scotter, 1994; Murphy & Shiarella, 1997; Organ & Ryan, 1995). George and Brief (1992) theorized about the importance of having a positive mood at work as the direct precursor to acts of organizational spontaneity like helping coworkers, defending the organization, making constructive suggestions, developing oneself, and "spreading goodwill." Citizenship could also result from cognitive processes regarding the norm of reciprocity and social exchange theory, (e.g., due to being satisfied with the job or expecting future rewards). Puffer (1987) found that a high need for achievement, high satisfaction with material rewards, and low perceived peer competition were related to more prosocial behaviors.

Organ and Ryan (1995) conducted the earliest, comprehensive review of dispositional and attitudinal predictors of OCBs. Conscientiousness produced small moderate validities with the altruism and generalized compliance dimensions of OCB (ρ = .22 and .30, respectively) after corrections for measurement unreliability. Agreeableness produced smaller validities with the same criteria (ρ = .13 and .11).

While citizenship performance has been linked to personality dimensions, and task performance has been linked to cognitive ability in research, I am not aware of any study that has examined the differential validity for personality and cognitive ability in predicting task and citizenship performance in a comprehensive review. Primary studies and meta-analyses of predictor relationships with one criterion or the other, however, suggest that the following predictions would be supported (Borman & Motowidlo, 1997; Motowidlo & Van Scotter, 1994; Organ & Ryan, 1995):

Hypothesis 3 (H3): Task performance will show a higher positive correlation with cognitive ability than with personality.

Hypothesis 4 (H4): Citizenship performance will show a higher positive correlation with personality than with cognitive ability.

In addition to the hypothesized predictors of cognitive ability and personality, there could be an enormous number of specific performance determinants that have unique relationships between task and citizenship performance. Thus, I/O psychologists have sought a subset of predictors that are manageable and carry utility, giving the "most bang for the buck" in understanding work processes. As a result, two sets of predictors are analyzed here. The first set consists of construct level measures, cognitive ability and personality. The second consists of two general methods commonly used in personnel selection (Muchinsky, 1997): structured interviews and biodata. Though it may seem odd to compare construct level measures with amalgamated measures, these comparisons are meaningful in light of the way that selection tools are administered; they are often mixed together. It is also acknowledged that some potentially powerful determinants of performance are excluded because meta-analysis requires that a sufficiently large body of

literature exists before it can provide accurate estimates. The next sections describe the predictor-performance relationships in more detail than Hypotheses 3 and 4 indicate. *Cognitive ability*

For decades, empirical evidence has accumulated to uphold cognitive ability as the single most consistent and strongest predictor of overall job performance, across a wide range of job types and situations (Gottfredson, 1997; Hough & Oswald, 2000; Hunter & Hunter, 1984; Jensen, 1998; Lubinski, 2000; Neisser et al., 1996; Schmidt, 2002; Schmidt et al., 1985). For many jobs cognitive ability appears to have a mean validity somewhere between .4 to .5 with overall job performance, after correcting for range restriction and criterion unreliability (Hunter & Hunter, 1984; Mayberry & Carey, 1997; Outtz, 2002; Ree & Carretta, 2002). This evidence is so strong, in fact, that general cognitive ability (GCA) is said to have validity generalization (Murphy, 2002; Ree & Carretta, 2002; Schmidt, Hunter, Pearlman, & Shane, 1979; Schmidt et al., 1985; Viswesvaran & Ones, 2002), meaning that a large percentage "of all values in the distribution [across jobs on which generalization evidence is based] lie above the minimum useful level of validity" (Schmidt, Hunter, McKenzie, & Muldrow, 1979, p. 618).

Some have even concluded that task dimensions within the same job are unlikely to moderate GCA test validities after correcting for artifactual variance in the distribution of validity coefficients observed in research (Schmidt & Hunter, 1977; Schmidt, Hunter, McKenzie et al., 1979; Schmidt, Hunter, Pearlman et al., 1979; Schmidt, Law, Hunter, Rothstein, Pearlman, & McDaniel, 1993), where artifacts account for up to 87% of the variance on average (Schmidt et al., 1993). "Only a measure of overall job performance

is needed in validity studies" when the corrected criterion reliability is high (Schmidt, Hunter, & Pearlman, 1981, p. 175). While statements like this imply that there is no need to differentiate between performance criteria for making practical selection decisions (Schmidt et al., 1981; Schmidt et al., 1985, Viswesvaran & Ones, 2002), there exist other reasons for examining performance relationships at a finer level of detail, some of which have been mentioned above in a more general context.

First, the use of global performance ratings obscures the meaning of one-to-one causal relationships between cognitive ability and different performance behaviors. Though it may be the case that some general ability, such as 'g,' allows some individuals to excel at virtually everything and others to be generally limited, furthering the theoretical understanding of job performance and its nomological network depends on examinations of why performance behaviors are linked to each other and caused by the same or different antecedents.

Second, the homogeneity of effect sizes may be just the result of various biases that are known to affect overall performance ratings including halo and leniency (e.g., Murphy & DeShon, 2000; Solomonson & Lance, 1997), personal attraction and racial bias (e.g., Ford et al., 1986; Pulakos, White, Oppler, & Borman, 1989), assimilation and contrast effects (e.g., Kravitz & Balzer, 1992), and the setting in which performers are observed by others (Rothstein, 1990). The reliability estimate for an overall rating might then be artificially high even though the measure fails to capture "true scores."

Third, the estimates of reliability (e.g., .60 in Schmidt & Hunter, 1977) that have been used to correct the validities of cognitive predictors (Schmidt et al., 1993; Schmidt et al., 1985) could be underestimates of true reliability if the observed variance in

performance scores is actually due to its dimensional nature rather than to error. Reliability estimates that are too small will result in overcorrections and will eliminate true variance (Algera, Jansen, Roe, & Vijn, 1984; Guion, 1998; Murphy, 1997), leading to the dubious conclusion that relationships are not moderated. Interestingly, Hunter and others (Schmidt, Hunter, McKenzie et al., 1979; Schmidt, Hunter, Pearlman et al., 1979) listed criterion contamination and deficiency as one of the seven likely sources of artifactual variance but have always considered it statistically uncorrectable. This study addresses this notion, in some sense, by manually separating different types of criteria based on their overlap with parts of the job performance construct domain.

Fourth, cognitive predictors almost always leave additional variance in overall performance unexplained (Schmidt, 2002; Schmidt et al., 1985; Schmidt & Hunter, 1998; Viswesvaran & Ones, 2002). If this unexplained part of performance consists of the same stuff, then it might be considered a separate dimension of performance that is not predictable by cognitive ability.

It is important to recognize that none of the above reasons have been thoroughly tested and that none disproves the claim of situational specificity for validity generalization (i.e., cognitive ability has a single true validity that does not vary across situations). The four reasons presented merely provide some compelling justifications for further investigating the relationship between cognitive ability and dimensions of performance. Any evidence found to support the distinction between task and citizenship performance would have meaningful implications. Cognitive test validities would be expected to show less variance and may be higher than previously expected if cognitive ability has a stronger link specifically with task performance, as theorized. Also, the

adverse impact on racial minority groups in selection that tends to result from using cognitive tests can be reduced by focusing on a dimension of performance that is less related to cognitive ability, assuming that dimension is predicted well by other variables (e.g., Hattrup, Rock, & Scalia, 1997).

Some prior research distinguishing dimensions of performance in validation research provides expectations for this study. In a small-scale "meta-analysis" of three studies (including Project A, an army selection study) with widely discrepant findings, Hattrup and others (1997) estimated cognitive ability to be correlated .41 with task performance and .16 with contextual performance when corrected for unreliability and range restriction. Murphy and Shiarella (1997), however, used another set of studies (also including Project A) to estimate the same relationships as .50 and .30, respectively.

The wealth of validation research on cognitive ability has also provided convincing evidence that job complexity moderates relationships between cognitive ability and overall performance (Hunter & Hunter, 1984; Murphy, 2002; Ree & Carretta, 2002). Validities with overall job performance have been estimated to be .58 for professional-managerial jobs, .56 for high-level, complex technical jobs, .51 for medium complexity jobs, .40 for semi-skilled jobs, and .23 for completely unskilled jobs (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). This may occur because cognition is partly defined as the ability to deal with complex situations (Gottfredson, 1997). Therefore, it is believed that:

Hypothesis 5 (H5): Job complexity will moderate the relationship between cognitive ability and task performance.

This hypothesis is essentially a replication of past meta-analyses, although this moderating effect is not of focal interest.

Job complexity may also be confounded with a greater need for interpersonal interactions and citizenship-like behaviors in higher level jobs, as discussed earlier and stated in H1 (Latham & Skarlicki, 1995; Gottfredson, 1997; Murphy & Cleveland, 1995). Thus, job complexity may moderate validities for at least two reasons: 1) high cognition allows workers to deal with complexity in tasks or the job or 2) citizenship performance is weighted more in complex jobs due to the social nature of the work. If H1 is supported (i.e., citizenship performance is more strongly related to noncognitive constructs in managerial and sales jobs) and H5 is not supported, it could be the case that job complexity moderates overall performance through aspects of citizenship rather than task performance; the job complexity and managerial distinctions may act similarly and be functionally equivalent. Finally, there appears to be no *a priori* reason for hypothesizing that the relationship between citizenship performance and cognitive ability will be moderated by job complexity, especially given Borman and Motowidlo's (1997) assertion that citizenship behaviors are common to all jobs.

Hypothesis 6 (H6): The relationships between cognitive ability and citizenship performance will be stable across job types based on complexity.

Personality

Though personality research has played an on-and-off role in I/O research, earlier studies lacked adequate theoretical frameworks (e.g., Guion & Gottier, 1965). It was not until after Barrick and Mount's (1991) seminal meta-analysis of Big Five (Conscientiousness, Agreeableness, Extraversion, Emotional Stability, and Openness to Experience) dimensions' validities with job proficiency that personality research regained popularity as a job performance predictor (Hurtz & Donovan, 2000; Salgado, 1998). A considerable body of research has emerged since then to support a clear link between some of the Big Five dimensions and certain aspects of job performance (Barrick, Mount, & Judge, 2001; Borman et al., 2001). Findings from a few studies are described briefly to provide an approximate idea of the relationships relevant to this study.

Four of the Big Five dimensions appear to be weakly related to cognitive ability, producing correlations of less than .10 (Bobko, Roth, & Potosky, 1999; Boudreau, Boswell, Judge, & Bretz, 2001; Cortina, Goldstein, Payne, Davison, & Gilliland, 2000). As would be expected openness to experience, or intellectance, has been shown to correlate moderately (r = .21 in Boudreau et al., 2001).

In Project A, conscientiousness correlated .11 with core technical proficiency (task performance), .09 with general soldiering proficiency, .22 with effort and leadership, and .30 with personal discipline (McHenry et al., 1990). Emotional stability was correlated .10, .12, .19., and .11 with those same variables, respectively. Conscientiousness and emotional stability correlated .32 with each other.

One study (McManus & Kelly, 1999) found that task performance was significantly related to extraversion (r = .22), and citizenship performance was related to extraversion (r = .29), agreeableness (r = .20), emotional stability (r = .23), and openness to experience (r = .23). Another study (Beaty et al., 2001) found emotional stability to be significantly correlated with task and contextual performance (.36 and .31, respectively). Emotional stability significantly correlated with task performance as well (.24).

Organ and Ryan (1995) provided perhaps the best, most comprehensive metaanalysis of the relationship between personality predictors and citizenship performance facets. Conscientiousness produced the largest correlations (.21 and .30, corrected) with different aspects of citizenship. Borman and others (2001) updated these findings with 20 additional studies and found citizenship facets to have uncorrected mean correlations of .24 with conscientiousness, .13 with agreeableness, and .08 with extroversion.

Hurtz and Donovan (2000) also meta-analyzed correlations between the Big Five and job performance, including measures of task and citizenship behaviors. All of the dimensions were weakly correlated with task performance and job dedication, with conscientiousness having the largest corrected mean correlations (.16 and .20, respectively). Emotional stability displayed the next largest relationships with task performance (.14), job dedication (.14), and interpersonal facilitation (.17) while the correlations for other personality dimensions averaged below .10. However, the credibility intervals indicated stable validity estimates only for the criterion of interpersonal facilitation.

Overall, it is clear that conscientiousness has consistently produced the strongest uncorrected validities with task performance (with a true r of about .20) and citizenship performance (about .24). Emotional stability has also produced moderate correlations with citizenship performance on a somewhat inconsistent basis, and low but significant correlations with task performance. Extroversion, on the other hand, has typically shown weak correlations with both performance dimensions. Agreeableness seems to have produced the most inconsistent results. Finally, openness to experience appears to be under-researched, which may result partly from studies showing its low correlation with

overall performance (Barrick & Mount, 1991). Based on theory and past findings, I pose the following hypotheses, where version A applies to citizenship performance and version B applies to interpersonal and job dedication facets. These apply only if H2 is supported:

Hypothesis 7 (H7A): Conscientiousness will be positively correlated with task performance and citizenship performance.

Hypothesis 7 (H7B): Conscientiousness will show a higher positive correlation with job dedication than with interpersonal facilitation.

Hypothesis 8 (H8A): Emotional stability will show a higher positive correlation with citizenship performance than with task performance.

Hypothesis 8 (H8B): Emotional stability will show a higher positive correlation with interpersonal facilitation than with job dedication.

Hypothesis 9 (H9): Agreeableness will be positively correlated with citizenship performance only.

Hypothesis 10 (H10): Openness to experience will be positively correlated with task performance only.

Structured interviews

Employment interviews, also proven predictors of job performance, generally contain ambiguity regarding the constructs being assessed (Bobko et al., 1999; Campion, Palmer, & Campion, 1997; McDaniel, Whetzel, Schmidt, & Maurer, 1994). Cognitive ability seems to account for less than 20% of the variance in interview ratings (Huffcutt, Roth, & McDaniel, 1996). Huffcutt, Conway, Roth, and Stone (2001) provided a framework of the constructs that are typically assessed in interviews to "provide greater

insight into why formats such as the situational interview predict performance and [to] allow interviews to be optimally designed to achieve specific outcomes such as high incremental validity and minimal impact on protected groups" (p. 897). They found that interviews in the research literature have primarily assessed basic personality constructs (35%), applied social skills (28%), mental capability (16%), and knowledge and skills (10%).

Unstructured interviews have no fixed format or set of questions and typically result in an overall rating for each applicant; structured interviews are the opposite (Schmidt & Hunter, 1998). Since this meta-analysis intends to distinguish predictor relationships with task and citizenship performance, unstructured interviews are unlikely to provide useful information and will be excluded. One implication of this restriction on study conclusions results from the fact that highly-structured interviews tend to assess applied mental skills and knowledge more than low-structured ones (Huffcutt et al., 2001). Thus, findings will not generalize to interviews with dissimilar content and structure.

Structured interviews have produced correlations in the range of .16 to .32 with aspects of task performance (Borman, 1982; Campbell, Prien, & Brailey, 1960). For a large sample of Air Force personnel, interviews produced median correlations of around .23 (ranging from .02 to .38) with a hands-on work sample of task performance and of global technical performance. When designed to predict dimensions of OCBs, situational but not patterned interviews were significantly correlated with job performance (.50 for OCB-O and .30 for OCB-I) (Latham & Skarlicki, 1995). Correlations of structured

interviews appear to be around .25 with cognitive ability and range from .12 to .26 with conscientiousness (Bobko et al., 1999; Cortina et al., 2000).

Based on these findings, structured interviews are assessing individual characteristics that are not limited to just general cognitive ability. These "other" skills, abilities, and motivation are likely to be useful for predicting citizenship performance, while knowledge and cognitive components are likely to predict task performance.

Therefore:

Hypothesis 11 (H11A): Interviews that primarily assess cognitive ability will be positively correlated with task performance.

Hypothesis 11 (H11B): Interviews that primarily assess personality will be positively correlated with citizenship performance.

Hypothesis (H11C): Interviews that assess both cognitive and personality constructs in approximately the same proportion will correlate positively with both task and citizenship performance.

Biodata

Biographical data, or biodata, have been shown to predict job performance relatively well and to exhibit smaller differences by racial subgroups than cognitive ability (Schmidt & Hunter, 1998; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Biodata are questions about past experiences that are somehow related to a criterion based on the premise that past behavior predicts future behavior (Mumford & Owens, 1987). They may be compound measures that comprise a range of constructs depending on what types of behavior are referred to in the questions (Mitchell, 1994; Nickels, 1994; Schmidt & Hunter, 1998). More recent biodata forms have also begun to include questions referring to past attitudes and values, in the hope that they predict future behavior – presumably mediated through future attitudes and values. It is also interesting to note that biodata are most akin to the most commonly used selection tools like résumés and job applications.

Correlations of biodata with cognitive ability have ranged widely from .05 to .50 (Bobko et al., 1999; Schmidt, 1988; Vinchur et al., 1998). In predicting overall job performance, empirical evidence has shown that biodata have minimal incremental validity over general cognitive ability (Schmidt & Hunter, 1998). This can result from the relationship between biodata and cognitive ability; it may also result if overall performance measures tend to exclude citizenship behaviors (again, a criterion problem rather than a predictor one).

Four biodata scales, including two dimensions of personality, in Project A had validities of .26 with core technical proficiency, .25 with general soldiering proficiency, .24 with effort and leadership, and .32 with personal discipline (Peterson et al., 1990; McHenry et al., 1990), where estimates are corrected for range restriction. McManus and Kelly (1999) showed that their biodata instrument for insurance sales representatives had similar relationships with contextual (r = .25) and sales task performance (r = .26).

Biodata appears to correlate moderately with cognitive ability (correlations between .05 and .27) (Bobko et al., 1999) and weakly (correlations near zero) with personality (McManus & Kelly, 1999). The relationship between biodata and structured interviews has been estimated to be between .08 and .27 (Bobko et al., 1999).

In conclusion, biodata have shown moderate correlations with both task and citizenship aspects of job performance behaviors, depending on what the specific

questions are designed to measure. They also appear to capture something other than personality which may still be a determinant of citizenship performance. Thus:

Hypothesis 12 (H12A): Biodata that primarily assess cognitive ability will be positively correlated with task performance.

Hypothesis 12 (H12B): Biodata that primarily assess personality will be positively correlated with citizenship performance.

Hypothesis 12 (H12C): Biodata that assess both cognitive and personality constructs in approximately the same proportion will correlate positively with both task and citizenship performance.

Summary of Research Hypotheses

H1: Citizenship performance will show higher positive correlations with noncognitive predictors in managerial and sales jobs than in other jobs.

H2: Effect sizes for measures related to citizenship performance will be moderated by the degree to which interpersonal facilitation and job dedication aspects are measured.

H3: Task performance will show a higher positive correlation with cognitive ability than with personality.

H4: Citizenship performance will show a higher positive correlation with personality than with cognitive ability.

H5: Job complexity will moderate the relationship between cognitive ability and task performance.

H6: The relationships between cognitive ability and citizenship performance will be stable across job types, based on complexity.

H7A: Conscientiousness will be positively correlated with task performance and citizenship performance.

H7B*: Conscientiousness will show a higher positive correlation with job dedication than with interpersonal facilitation.

H8A: Emotional stability will show a higher positive correlation with citizenship performance than with task performance.

H8B*: Emotional stability will show a higher positive correlation with interpersonal facilitation than with job dedication.

H9: Agreeableness will be positively correlated with citizenship performance only.

H10: Openness to experience will be positively correlated with task performance only.

H11A: Interviews that primarily assess cognitive ability will be positively correlated with task performance.

H11B: Interviews that primarily assess personality will be positively correlated with citizenship performance.

H11C: Interviews that assess both cognitive and personality constructs in approximately the same proportion will correlate positively with both task and citizenship performance. H12A: Biodata that primarily assess cognitive ability will be positively correlated with task performance.

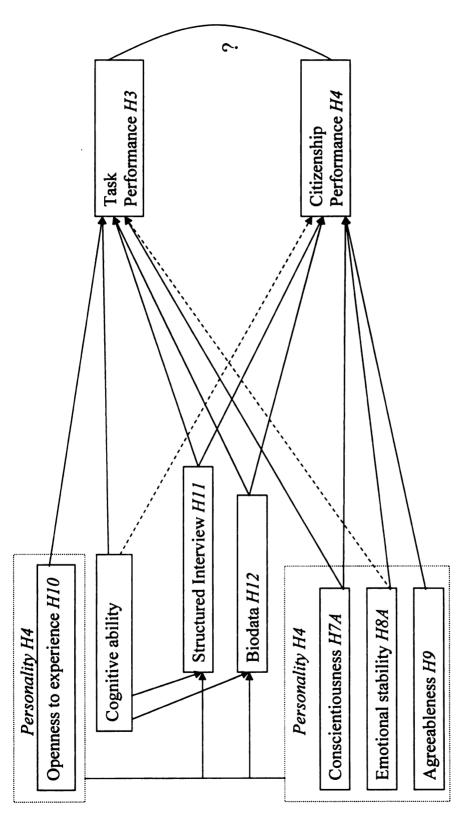
H12B: Biodata that primarily assess personality will be positively correlated with citizenship performance.

H12C: Biodata that assess both cognitive and personality constructs in approximately the same proportion will correlate positively with both task and citizenship performance.

*The hypotheses with asterisks are only relevant if support is found for H2.

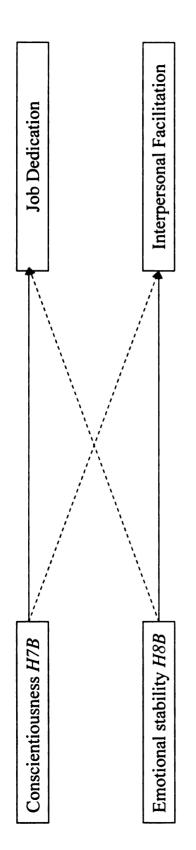
Figures 1 and 2 provide an integrated visual representation of these hypotheses, many of which are based on findings in the literature reviewed.

Figure 1. Model of Hypothesized Relationships Between Predictor Variables and Job Performance Dimensions.



Note. Dashed lines represent predictor-criterion relationships that are weaker than the solid lines from the same predictor. Hypotheses associated with each variable are italicized. Moderator hypotheses are not depicted.





Note. Dashed lines represent predictor-criterion relationships that are weaker than the solid lines from the same predictor. Hypotheses associated with each variable are italicized.

Conclusion

There is a large body of research on the validity of various selection tools. Cognitive ability is known to produce large validities, and this finding is generalizable across different types of jobs. Other individual difference characteristics have also produced respectable validities but in a less consistent manner. Because job performance is the outcome that many researchers try to predict and because it is a complex construct, there is a need to investigate more direct relationships between individual differences and different types of performance behaviors.

Recent advancements in theory and the development of explicit job performance taxonomies have stimulated research, particularly in the area of citizenship performance. As more studies have become available and theories have become more complex, it is appropriate to evaluate the past findings and determine what we know and what we do not know.

Meta-analysis is one useful method of summarizing and evaluating the information provided by many studies. By quantitatively cumulating past results, one can extract the effects of sampling error variance to which primary studies are bound, and derive better estimates of conceptual relationships. Meta-analysis also permits examinations of moderating influences on the distribution of observed validities that might not have been analyzed within any one study (Rothstein, McDaniel, & Borenstein, 2002). It is hoped that the summarizing discrepant estimates of interesting relationships (e.g., task and citizenship performance correlations) and detecting or supporting the existence of plausible moderators will refine estimates of relationships, by reducing unexplainable variance in results across studies, which may also allow us to expect higher

validities for some measures in particular settings or for particular groups (Barrick et al.,

2001; Salgado, 1998).

METHOD

Literature Search

I conducted a literature search using both computer-guided and manual approaches to find relevant published works between the years of 1988 (when "citizenship behavior" terms were defined) and 2004. The American Psychological Association's PsycINFO, PsycFIRST, ERIC, ABI/INFORM, and BusinessOrgs databases were used in the computer searches, including these keywords: citizenship, contextual performance, task performance, task proficiency, prosocial behaviors, structured interview, biodata, and biographical data. The manual search covered three prominent journals in the field of I/O psychology: *Journal of Applied Psychology, Personnel Psychology*, and *Organizational Behavior and Human Decision Processes*. The reference sections of seminal articles and meta-analyses were also used to locate additional studies. These methods together yielded 589 references to studies possibly containing codable information. The large yield was partly due to the large number of variables included in searches and partly due to erring on the side of inclusion when reading unclear abstracts. *Criteria for Study Inclusion*

Only a portion of the studies identified by the literature search yielded usable information. Studies were coded and analyzed if they met the following criteria: 1) they were written or translated in English, 2) there was a measure of the relationship between at least two of the study variables (i.e., task performance, citizenship performance, cognitive ability, the Big Five, biodata, interviews, but not overall performance¹), 3) individuals were the unit of analysis, 4) the statistical information necessary for

¹ Information about overall performance was included as supplemental information only when an effect size for at least one of the performance dimensions was reported.

computing a correlational effect size was presented.

Bobko et al. (1999) explained the importance of choosing appropriate studies for inclusion in a meta-analysis and of consistently applying the same decision rules. It is ideal to combine studies primary studies and to exclude meta-analytic estimates that were derived with unique decision rules (cf. Wanous, Sullivan, & Malinak, 1989). Then again, information is lost when meta-analyses contain studies that are not otherwise obtainable or eligible (e.g., studies before 1988 in this case). Thus, uncorrected cumulative correlations from meta-analyses were included in these analyses when they followed procedures similar to the one used here. As the performance taxonomy in this study has not been used in prior meta-analyses, certain studies (e.g., Barrick and Mount's 1991 meta-analysis) did not fall within the inclusion criteria either because they classified measures differently, particularly with respect to citizenship criteria. To avoid doublecounting studies and giving certain findings too much weight, meta-analyses were excluded when most of their primary studies were already included in this database (e.g., Borman et al., 2001), and primary studies were excluded if they were contained in a meta-analysis that otherwise provided a large amount of unique data.

Because laboratory research is often meant to generalize to work settings, studies including university students and people in non-work settings were included when the measure of performance mirrored actual job performance in some way. So, solely academic performance measures (e.g., GPA) were excluded.

Data Coding Procedure

A two-stage coding process was used to obtain information necessary for testing the hypotheses. In the first stage, the author and an advanced undergraduate in

psychology coded sample characteristics based on information from studies identified through the literature search that met the criteria for inclusion. In the second stage, characteristics related to the hypothesized moderators were coded by multiple raters knowledgeable about I/O psychology. Concurrently, the author used the O*NET database (http://online.onetcenter.org) to assign job complexity codes based on the *Dictionary of Occupational Titles* to test Hypotheses 5 and 6. The two stages are described in more detail below.

For the first stage, a pilot coding sheet was developed based on guidelines and examples from Lipsey and Wilson (2001). Items necessary for analyzing the specific hypotheses were added. The pilot coding sheet (Appendix A) was then sent to a college graduate outside of psychology to determine where the instructions were too vague, ambiguous, or confusing. As a result, the coding sheet was simplified and made shorter.

The resulting coding sheet (Appendix B) was then used in subsequent phases of coding, allowing data to be recorded for each sample within a study in a new spreadsheet. Items on the coding sheet referred to the number of independent samples in the study, sample sizes, a qualitative description of the sample, whether a sample consisted of managerial employees, whether the study design was predictive or concurrent, reasons for missing data, whether a manipulation occurred between measures (as in training studies), a description of each measure, a judgment about measure subjectivity/objectivity, a judgment about measure broadness/narrowness, measure reliability, and correlational effect sizes between variables qualifying for this meta-analysis based on the definitions in Table 1.

The undergraduate student had taken courses in research methods and in statistics, and underwent training in how to code studies according to a code book of rules, including variable definitions. After the training over six weekly meetings, the undergraduate coded four studies for comparison with the author's coding of the same studies. The coders met and discussed disagreements and ambiguity about the coding rules and operational definitions of variables. The code book was revised based on these discussions, and the revision was used to code the remaining studies. The operational definitions eventually used are those in Table 1 and the coding rules are presented in the code book (Appendix C).

Interrater Agreement

Both coders examined a subset of studies and agreed on 56 of 58 (96.6%) about which studies provided codable data. For these studies, the coding sheet and code book were used to collect study characteristics, measurement characteristics, and correlational effect sizes, as described earlier. Information about interrater agreement is provided in Table 2 for 29 studies that were deemed codable. For categorical variables, I computed kappa to index the level of agreement achieved beyond chance agreement (Table 2 and Appendix D). Generally, kappa values of .8 or higher are very satisfactory, between .6 and .8 are good, between .4 and .6 moderate, and of less .4 poor (Landis & Koch, 1977). The kappa values for coded information indicate that moderate agreement existed for judgments about whether samples were managerial ($\kappa = .58$) and whether measures were broad ($\kappa = .52$). Kappa values between raters regarding the study design and whether measures were subjective were lower.

Table 1

Study variable labels and definitions.

Variable	Definition		
Task performance	Behaviors that directly or indirectly (through other workers) affect core production that transforms input to output or delivers a service. Sometimes these are referred to as "in-role" behaviors because they are tied to one's job roles. However, the two concepts may be very different if the role includes non-task behaviors.		
Citizenship performance	Behaviors that 1) are not directly related to core tasks and 2) support the social and/or psychological environment. Examples include: loyalty, cooperative behaviors (not affecting core production), whistle-blowing, sportsmanship, prosocial behavior, personal initiative, showing extra effort and perseverance, volunteering to do extra or unrelated work. Counterproductive or retaliatory behaviors are NOT included.		
Job Dedication (Citizenship performance)	Citizenship behaviors that do not require a direct interaction with another person. Instead, they are related to helping the organization overall. Examples: working hard, taking initiative, and following organizational rules.		
Interpersonal facilitation (Citizenship performance)	Citizenship behaviors that require a direct (not necessarily face to-face) interaction with another person. Examples: helping others and backing people up.		

Table 1 (cont.)

Variable	Definition				
Cognitive ability	Broadly, any computational, problem solving, or mental abilities.				
Personality	Enduring, characteristics of the individual that are tied to one of the Big Five dimensions: Conscientiousness, Agreeableness, Extroversion, Emotional Stability, and Openness to Experience. A measure may capture smaller aspects of any one dimension but not overlap with another dimension.				
Conscientiousness	dependability, achievement striving, and planfulness				
Extraversion	sociability, dominance, ambition, positive emotionality, and excitement-seeking				
Agreeableness	cooperation, trustfulness, compliance, and affability				
Openness to experience	intellectance, creativity, unconventionality, and broad- mindedness				
Emotional stability	lack of anxiety, hostility, depression and persona insecurity				
Structured interview	A structured interview, at the very least, evaluates a response to each question posed to the interviewee (from Huffcutt & Roth, 1998)				
Biodata	A measure of background life experiences that is intended to predict future behaviors of the same type				

Note. The structured interview and biodata are predictor measures but the other variables refer to constructs (and appropriate measures). Also, simple demographics were not treated as biodata, per the definition given here.

For continuous variables such as the sample size of a study (N), percentage agreement is reported in Table 2. Most of these are nearly 90%. For sample size, agreement for the exact N recorded was the same 83% of the time but rarely differed by more than 5 cases. The discrepancies often resulted from differences between the Nstated in the sample description and the N reported for a correlation matrix after removing some unusable cases. The lowest agreement occurred for the categorization variable that was formed; the creation of this variable and its meaning is explained below.

The estimates above are imperfect because some of the codes are dependent on each other (i.e., a miscode in one place will cause a subsequent miscode). Some attempt was made to evaluate this effect by treating certain disagreements as categorization errors when a measure was not labeled the same way between raters but all other information pertaining to that measure was correct. Codes that were different only because of an earlier categorization error were not treated as a disagreement but the categorization error itself was tallied. For example, one rater treated a set of supervisory ratings as task performance while the other rater treated them as overall performance but both raters coded the reliability estimate for that measure accurately from the original study. This was counted as a categorization error but not an error in recording the reliability estimate. The percentage of categorization agreement is shown in Table 2 and represents the times that raters classified measures from primary studies into the same study variables used here. The percentage provides some indication of how generalizable the coding scheme in this study would be if applied in other meta-analyses.

The percentage is fairly low but is an underestimate of the coding scheme's reliability for two specific reasons, apart from the general inexperience of the raters in

conducting meta-analyses. First, the classification of certain variables determined how later variables could be classified (e.g., a job dedication variable incorrectly coded as task performance could not be coded correctly as either citizenship or job dedication). This phenomenon could not be evaluated because one categorization error could lead to one or more other categorization errors depending on the actual disagreement. Second, the percentage does not include agreement on decisions of omission when both raters decided not to include certain variables from primary studies. The percentage agreement would rise for every variable that both raters decided was too different from the definitions in Table 1 to be included. Additionally, it should be stated that virtually no estimate of interrater reliability is free from bias; two raters could demonstrate excellent consistency but be "wrong" if they make the same errors.

Specifically with regards to the percentage agreement for r (i.e., how many times raters recorded numerical correlation values in the same way), the percentage obtained was good but not as high as one would expect, given that numbers are listed in tables and errors are less related to differences in judgment between raters than to transcription problems. However, a substantial percentage of the coding disagreements were due to the undergraduate providing the wrong sign for "neuroticism" correlations in four studies, as opposed to its positive direction, "emotional stability." That is, the correlations were coded as the same variable and were of the same magnitude but of opposite signs. If these trivial errors are excluded since they can most likely be eliminated with additional training, agreement rises to 93%.

Table 2

Interrater Agreement of Coded Data.

Kappa Agreement							
Managerial	Design	Broad	Subjective				
<i>K</i> =58	<i>K</i> =29	<i>K</i> =52	<i>K</i> =43				

Percentage of Agreement								
# of Samples	N	Categorize	Reliability	Туре	<u>r</u>			
89.3%	83.3%	73.6%	89.5%	89%	86.9%			

Note. Managerial = 3-choice item about whether the sample included managers, nonmanagers, or both. Design = study design (predictive or concurrent). Broad = subjective assessment about criterion relevance of each measure. Subjective = subjectivity of each measure. # of samples = number of independent samples in the study for which data was coded. N = study sample size. Categorize = times correlations were associated with the same variable labels. Reliability = numerical estimate. Type = reliability index used. r = numerical correlations between variables in primary study.

For these 29 studies, disagreements were discussed between the coders and resolved. One study was excluded because it was not clear whether the performance variable fit cleanly into any of the categories. Due to resource constraints and because the key variables showed relatively good agreement, I coded the remaining studies. The Design, Broad, and Subjective codes were not used in later analyses due to low agreement and because they are not related to the hypotheses. They were coded as a precautionary measure for using post hoc tests to explain strange cases.

To enhance the reliability and generalizability of the classification decisions, previous sources of literature were used during the remainder of the coding process. I used John's (1990) "Big Five" taxonomy that maps the subfacets of popular measures such as the Jackson PRF, NEO-PI, and Hogan PI into the five categories. When reasonable, I used coding rules from previous meta-analyses that contained data at the level of the primary study (e.g., Cortina et al., 2000). However, previous work was not relied on when different conceptual definitions were used or when overall job performance was the sole criterion.

For the second stage of coding, the author used the O*NET online database to obtain a rating of job complexity for jobs included in primary studies. It provides the "specific vocational preparation (SVP) range" from the *Dictionary of Occupational Titles* (DOT) (U.S. Department of Labor, 1991). SVP is defined as "the amount of lapsed time required by a typical worker to learn the techniques, acquire the information, and develop the facility needed for average performance in a specific job-worker situation" (Appendix C of the DOT) and is measured with a 9-point scale² ranging from "Short demonstration" to "Over 10 years." Because O*NET only references actual jobs, this procedure was

carried out for actual jobs in the primary studies (unless a simulated laboratory task closely resembled a real job). For mixed samples, a lower SVP value was used if the jobs did not differ by more than one point.

Finally in the third stage of coding, biodata and interviews were to be coded according to the constructs being assessed by them. In the end, only biodata was deemed usable as too few studies provided correlations related to structured interviews. The author and three graduate students in I/O psychology who were familiar with terms and definitions but blind to the study hypotheses provided ratings of the biodata measures used in 15 studies that reported correlations related to Hypothesis 12 (i.e., with task or citizenship). Raters decided what percentage of the measure assessed the hypothesized predictors: cognitive ability and personality (excluding extraversion because it was not hypothesized to have a significant relationship to either task or citizenship performance). Raters were told that the percentages did not need to sum to 100 and that all other things being assessed in the biodata should be attributed to the remaining percentage.

Because Hypothesis 12 makes a higher level distinction between biodata that primarily assess cognitive ability, primarily personality, or both, the information provided by raters was recoded to fit the broader categories rather than specific personality dimensions. That is, the percentages for the four personality dimensions were aggregated into a composite percentage representing the amount of biodata assessing personality overall. This resulted in a percentage estimate of cognitive ability and of personality assessed by biodata provided by each rater. Across the four raters and two percentage values (cognitive and personality) for all studies, interrater reliability (i.e., the intraclass

² Often, a range was given (e.g., "below 4"). The lowest possible number was used.

correlation) was .89, and all raters agreed that one study did not provide enough information to be coded. This level of reliability was determined to be very acceptable since agreement between coders of subjective variables like methodological quality tends to produce low agreement (Hattie & Hansford, 1984; Viswesvaran, Ones, & Schmidt 1996).

In addition to the planned coding processes described above, I coded all studies for range restriction based on the sample setting. The "study design" variable was originally coded to provide an index of range restriction for applied samples but it was too confusing because measures were often administered at different times even though the design was concurrent, it had a low interrater agreement, and it did not necessarily mean that a given measure was, in fact, range restricted. A new code was assigned to each study (1 =restricted range, 0 =unrestricted range) based on whether a sample's participants were explicitly selected to meet some threshold level (directly or indirectly) on one of the study variables (i.e., either predictors or job performance). Although this decision was subjective, the end result essentially was that only samples drawn from the general public or job applicants were counted as unrestricted, making the code fairly clear. Job incumbents, successful performers, and college students were examples of samples considered to be restricted. Restriction here was viewed simply as an influence that would attenuate observed correlations representing the true relationship between variables.

Meta-analytic Procedure

I generally followed the strategy developed by Hunter and Schmidt (1990; with refinements detailed in Hunter & Schmidt, 2004) but supplemented analyses with the

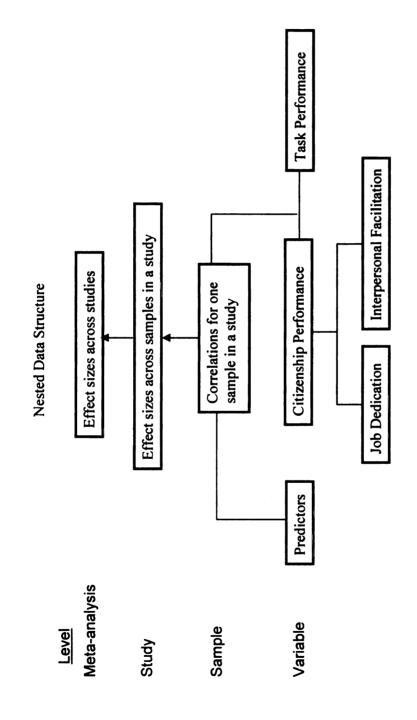
multivariate method proposed by Becker and colleagues (Becker, 1992; 1996; 2000; Kalaian & Raudenbush, 1996; Raudenbush, Becker, & Kalaian, 1988). The multivariate framework, as well as traditional decision rules, was used to manage multiple levels of dependencies in the data that would otherwise violate basic statistical assumptions. "Although all [primary study] design features lead to dependence among study outcomes. the nature of the dependence depends on exactly what comparisons are computed and the metric(s) in which they are expressed" (Becker, 2002, p. 501). At least 3 levels of dependence were evident in this database: multiple measures of the same subfacet of a study variable (e.g., altruism and courtesy formed the interpersonal facilitation variable, a facet of citizenship), the same study variable measured more than once or with separate facets (e.g., general cognitive ability measured with verbal and numerical tests), multiple correlational effect sizes between different variables within the same study (e.g., the correlation between cognitive ability and task performance, and biodata and task performance), and correlations reported for more than one sample (e.g., employees in two separate organizations). Figure 3 shows these levels of the data structure that contribute to meta-analytically derived estimates.

I began by using the Hunter and Schmidt (2004) criteria for cumulating findings within studies/samples (p.429-442). When multiple measures of the same variable (as defined in this study) were used, I computed a composite correlation using equation 5-8c in Nunnally and Berstein (1994) and a corresponding estimate of the composite's reliability (with Eq. 7-15) when sufficient information was available. Obviously, results that were believed to be distinct (i.e., task versus citizenship performance) were not combined.

Multiple groups sampled within the same study created a second level of dependence. Subgroups were treated as individual cases in the meta-analysis when their results could be thought of as fully replicated designs, since the results could have been published separately. (Whether this step achieves an acceptable level of independence is relative; unique studies are often treated as independent despite having been conducted by the same researchers with the same measures.) If the grouping variable was not related to a moderator in this study (e.g., race or gender), total group correlations were used, as recommended by Hunter & Schmidt (2004).

A third level of dependence existed because primary studies typically provided data for more than one relationship between variables relevant to this meta-analysis. Multivariate techniques for modeling dependencies between correlations within the same study were applied after correlations were corrected for artifactual variance. The specific procedure is described after the next section on statistical corrections. Ultimately, 196 studies provided usable data. Initial statistics for the mean sample-size weighted correlation of observed effect sizes were computed. Although many of the resulting statistics are important primarily for their use in subsequent computations, one should consider the mean weighted correlations and their confidence intervals. When the confidence interval does not include zero, the effect size is regarded as statistically significant (cf. Hedges & Pigott, 2001), as is the case with typical confidence intervals. According to Whitener (1990), a confidence interval "reflects the effect of sampling error and is therefore applied to sample-size weighted mean effect sizes that have not been corrected for research artifacts" (p. 316).

Figure 3. Multilevel Structure of Meta-analytic Database.



Corrections for Artifactual Variance

Artifacts can systematically alter the magnitude of observed relationships and inhibit theory testing by introducing additional variance in estimates cumulated across studies (Paese & Switzer, 1988; Viswesvaran et al., 1996). Hunter and Schmidt (2004, p. 76) describe 10 "potentially correctable" study artifacts that alter observed correlations, but the traditional practice in validation research has been to correct only for sampling error, measurement unreliability (typically in the criterion), and sometimes for range restriction (Hunter & Schmidt, 2004; Raju, Burke, Normand, & Langlois, 1991). Nonetheless, the appropriateness of any correction depends on how well it models the error that is assumed to be distorting observed results; the correction of artifactual variance is more of an art than a set procedure. In this study, I simultaneously corrected for two types of artifacts: the dichotomization of variables and measurement unreliability using formulas provided by Hunter and Schmidt (2004).

Measurement unreliability. While "a thorough investigation of the criterion domain ought to include an examination of the reliability of dimensions of job performance" (Viswesvaran et al., 1996, p. 557), there has been some debate about the accuracy and meaningfulness of different reliability indices (e.g., Murphy & DeShon, 2000; Schmidt, Viswesvaran, & Ones, 2000). Reliability estimates were most often provided by studies in the form of internal consistency (alpha) but sometimes test-retest or interrater correlations. Each form has advantages and limitations (Schmidt & Hunter, 1996; Viswesvaran et al., 1996) but none completely captures the concept of reliability and the use of any one inevitably leads to some amount of over- or under- corrections (Cortina, 1993; Cronbach & Shavelson, 2004; Hunter & Schmidt, 1990; Murphy &

Deshon, 2000). Thus, one must make "a best guess" about a measure's reliability based on reported information.

Alpha was most commonly reported in the studies included here but is an inappropriate reliability estimate when 1) distinct dimensions are rated (Nunnally & Bernstein, 1994) and/or when 2) a measure contains error from its items and from a rater's judgment (Hunter & Schmidt, 2004), as in supervisor ratings of performance based on a dimensional form. In the first situation, I used other estimates of reliability (e.g., test-retest correlations) and not alpha because it underestimates true reliability. In the second situation, however, I used alpha if estimates that are more appropriate were not available and the rating was not clearly dimensional. Alpha is an overestimate of true reliability in those situations and will cause correlations to be undercorrected for the influence of artifactual variance. While the estimated p will have more error, overall conclusions will be more conservative³ and be more likely to produce Type I errors than if some error variance was mistakenly attributed to measurement unreliability and removed. Also, it is better to correct a sample correlation partially than to make no correction because failing to apply a correction procedure consistently to all studies will contribute to variance between study estimates, confusing the results of moderator analyses.

Although an adequate description of the measurement method was not always provided, I attempted to use the most appropriate reliability index reported. For example, if a measure consisted of multiple dimensions and had low internal consistency but high

³ The mean meta-analytic correlation will be biased downward but will retain meaningful variance due to the multidimensionality of a measure along with the additional error variance.

test-retest reliability, I used the latter. When a single construct was measured with a set of written questions and internal consistency was low but interrater reliability was high, I used the internal consistency under the premise that raters cannot produce reliable true scores using an unreliable tool.

Next, I borrowed estimates of reliability from other published sources when a primary study did not report the observed reliability of a measure. For example, some estimates for different kinds of performance ratings were taken from the meta-analysis of reliability by Viswesvaran et al. (1996). Although full reliability information was not available in the end, there was enough to warrant individual corrections rather than the use of artifact distributions.

Finally, I used mean substitution to estimate reliability for measures when no relevant estimate could be located in the literature. Imputation was conducted so that corrections would be applied to the data uniformly. As long as the mean value is somewhat accurate, the remaining correlations will increase by the same amount as the correlations with reliability estimates reported, preserving observed variance (i.e., not creating additional variance by correcting some correlations and not others). The resultant set of reliability estimates was used to correct correlations for predictor and criterion unreliability in order to estimate the theoretical relationship between each of the study variables (Hunter & Schmidt, 2004; Orwin & Cordray, 1985; Salgado, 1998).

Range variation. Range restriction, or variation, across studies is also important to consider because it will alter the magnitude of correlations and can create artifactual variance in observed correlations. A variable measured within a limited range will attenuate the maximum possible correlation that can be obtained, and a variable measured

with a range larger than is found in the real world can inflate correlations. When studies are differentially affected by the effects of range variation, with some correlations being attenuated and others inflated, range variation will cause artifactual variance in the distribution of observed correlations. Despite these problems, corrections for range variation should only be made when one can accurately model the error created by differences in ranges across studies. If one cannot make proper assumptions about how true correlations are being distorted by range variation, it is impossible to remove the distortion accurately. I did not consider range variation to be correctable in this study for a number of reasons. In their meta-analysis, Organ and Ryan (1995) too felt that they could not accurately specify what a normal range of variation would be on OCB measures.

Range variation is often removed through the use of artifact distributions by identifying a common reference group to which all studies should be calibrated. In this database, studies came from diverse settings including the general public, university students, job applicants, and job incumbents for different kinds of organizations. As a result, I could not assume a useful hypothetical range for a common reference group, precluding the use of artifact distributions as the basis for range restriction corrections.⁴

Corrections for range variation could still be made on a case by case basis, but only when primary studies provide sufficient information about the range with respect to the variables measured (e.g., selection ratios or turnover rates). Few of the studies in this database included such information. Less than five studies included measures actually used in selection.

⁴ Sackett and Ostgaard (1994) found small differences between job-specific applicant pools and national samples. Yet, there is no immediate evidence that university applicants are similar to job applicants.

Still another method of correction is to use the variance associated with different variables as an indicator of range effects. Variance estimates for measures were provided fairly often in primary studies. However, this method of correction is only appropriate when studies use the same measures (cf. Raju, Pappas, Williams, 1989). This was not the case here. Furthermore, no method allows a clear way to estimate the combined effect of (direct and indirect) range variation on multiple correlations reported for different variables in the same study, especially when some relationships might be attenuated and others enhanced. For example, a study predicting the performance of job incumbents or successful workers may be range restricted on task performance but range enhanced on citizenship behaviors if citizenship is related to helping the organization by participating in the research study.

Although range restriction and other types of artifacts typically affect the magnitude and variation of effect sizes observed in research, sampling error appears to account for the bulk of artifactual variance, especially when the sample sizes in primary studies are small (Koslowsky & Sagie, 1994), accounting for more than 70% in some studies (Schmidt et al., 1993). In conclusion, this meta-analysis adopts the perspective that no correction is better than a poor one.

The correlations corrected for statistical artifacts were then cumulated across studies, weighted by sample size and the size of artifact correction, to produce an estimate of the population correlation and its variance. The estimate of variance was used to assess the presence of moderators. As all of the analyses here were conducted on less than 60 studies, the power and accuracy of most methods for identifying moderators are relatively low but comparable (Sagie & Koslowsky, 1993). But because no test is

definitive, I used three approaches, two of which are recommended by Hunter & Schmidt (2004).

First, I applied Hunter and Schmidt's (1990) "75% rule" (of thumb) to compute the percentage of variance explained by artifacts (i.e., error). However, I lowered the threshold to 60% as recommended by others (Colquitt, LePine, & Noe, 2000; Hom, Caranikas-Walker, Prussia, & Griffeth, 1992; Koslowsky & Sagie, 1994; Mathieu & Zajac, 1990) since correlations were not adjusted for range variation. When 60% of the observed variance is due to sampling error, measurement error, and variable dichotomization, evidence for the presence of a true moderator was judged to be small.

Second, I calculated an estimate of the true range of population correlations with credibility intervals. When these are large or overlapping 0, they too suggest the presence of moderators (Hunter & Schmidt, 1990; Tett & Meyer, 1993). Koslowsky and Sagie (1993) provide guidelines about how big an interval should be before it indicates the presence of a moderator: roughly larger than .11.

Third, I computed the Q statistic and its chi-square value to determine whether the amount of observed variance would be larger than that expected based on chance. This method also allows significance tests to be computed for hypothesized moderators. The Q values for meta-analyses conducted on subgroups based on a moderator are then compared to the total group Q with a chi-square test (where the degrees of freedom equals the number of subgroups minus 1). Together, these qualitative (Hunter & Schmidt, 2004) and quantitative (Cooper, 1998) comparisons allowed me to determine whether it was likely that moderators were present and whether hypothesized moderators explained observed variance in correlations.

Multivariate Meta-analysis Procedure

A single meta-analysis provides information about the relationship between two variables. However, researchers are often interested in examining the larger pattern of relationships between predictors and multiple outcomes that reflect a realistic phenomenon. The same is true for this study, and the relationships of key interest are for the multiple dimensions of performance. Unfortunately, a level of dependence is created in the data that most likely "affects Type I error levels in complicated ways" (Becker, 2000, p. 503) when multiple outcomes are reported by primary studies (e.g., correlations between predictors and task and citizenship performance). At least a few analytic approaches have been employed to deal with this dependence, some being more valid than others (Becker, 2000; Hunter & Schmidt, 2004).

One practice is to combine multiple correlations reported for the same sample into a single effect size, but at the cost of losing information. This practice is particularly limiting when distinct constructs are combined into a less meaningful, broader unit (e.g., overall performance), and was not a viable option here as the distinction between different criteria was the central research goal.

Others have conducted a series of meta-analyses for each pair of variables under consideration. A single meta-analysis will provide the best estimate of a correlation based on the primary studies cumulated, and unique multiple meta-analyses will provide good estimates of unique correlations. However, it is often the case that some primary studies report multiple correlational effect sizes. When these effect sizes are interrelated because they come from the same study, they are statistically dependent and provide less information than a set of unique, independent correlations. Thus, a study that contributes

correlations to a series of meta-analyses provides some redundant information, and this information about a study that is common across meta-analyses can produce correlated errors in more complex analyses of the data (e.g., linear regression), violating traditional assumptions of independence that increase the rate of Type I errors (Becker, 2000; Bliese, 2002; Gleser & Olkin, 1994; Kenny & Judd, 1996; Raudenbush et al., 1988). Therefore, one must account for dependence between the units of analysis in order to compare the magnitude of two correlations in two separate meta-analyses where some of the data overlaps because some studies contributed correlations to both meta-analyses.

As an aside, dependence between studies is less problematic if the variables studied in separate meta-analyses are unrelated (Becker, 2000). In this study, there is no prior expectation that the relationship between the relevant outcomes of task and citizenship performance will be small or large. If the two dimensions are substantially related, it will be necessary to consider the dependence between primary studies/samples before comparing various correlations in linear models. Because H3, H4, H11, and H12 specifically predict that certain variables will have higher validities with one performance dimension or the other, I attempted to model dependence in the database using a relatively new multivariate method of meta-analysis described by Becker and colleagues (Becker, 1992, 2000; Raudenbush et al., 1988).

This multivariate method of meta-analysis models the dependence between outcomes reported for the same sample by treating data structures as meta-analytic cases (i.e., units of analysis) rather than individual correlations. "A fully multivariate approach should provide justifiable tests of significance for more complex questions than can be addressed using the ad hoc or univariate approaches described above, and more accurate

probability statements for all tests conducted" (p. 505, Becker, 2000). The Becker method is quite versatile as it allows pooled correlations to be calculated even when different studies contribute different effect sizes (Raudenbush et al. 1988). So, the data structure comprising each case may look very different because any sample may contribute one correlation or an entire correlation matrix to the analysis.

The method, however, does require as input data regarding the covariance between every effect size reported in the primary study. In other words, a study reporting a ability-task performance correlation and a conscientiousness-task performance correlation would also need to provide the correlation between ability and conscientiousness. When studies did not provide this necessary information, they were either excluded from the analysis or retained by borrowing relevant estimates from other literature (Raudenbush et al., 1988) or mean imputed (Becker, 2000). I set an arbitrary cutoff that a sample would be retained if it required less than 20% of its correlations to be imputed. Therefore, samples must have contributed at least 6 correlations before they were even eligible for imputation.

After correlations were corrected for artifactual variance using the Hunter and Schmidt (2002) procedure, they were used as input in the multivariate analysis⁵. I used formulas provided by Becker (2000) to construct a vector of correlations for each study and a corresponding variance-covariance matrix modeling the interdependence between sampling error for these correlations. (It was necessary to add Hunter and Schmidt's (2004) correction factor to Becker's (2000) equations 4 through 6 for computing the variances and covariances between correlations reported in a study.) I then used the

generalized least squares, fixed-effects approach (Becker, 1992, Raudenbush et al., 1988) to compute a vector of mean correlations cumulated across samples. These estimates are averages of corrected correlations across samples, weighted by sample size and the sample's variance-covariance matrix for effect sizes. The analysis also produces the pooled variance-covariance matrix used in weighting. To test Hypothesis 12, I used additional formulas provided by Raudenbush et al. (1988) to predict variation in effect sizes using study level characteristics. These analyses were conducted in SAS/IML (SAS Institute, 2001), and sample syntax of the analyses is included in Appendix E. (The program for these analyses would not run in SAS with correlations of 0 so corrected mean correlations equal to 0, when using two decimal places, were set to 0.00001 for input into SAS.)

The resulting estimates of population correlations were used to fill in an estimated " ρ -matrix" (Table 10). The ρ -matrix represents the best estimate of each correlation between study variables, as a total set. While researchers have tested the overall fit of models based on meta-analytically derived matrices (e.g., Carr, Schmidt, Ford & DeShon, 200; Colquitt et al., 2000; Shaw, Wild, & Colquitt, 2003; Tett & Meyer, 1993), there are a number of conceptual issues that one must address to justify conclusions derived from these analyses (Viswesvaran & Ones, 1995). The most obvious problem concerns estimating error variance in the total sample correctly because there is no single value for sample size that applies to the entire matrix (even though some have argued for the use of the harmonic mean). Also, the information contained in a meta-analytic matrix is based on pairwise (deleted) correlations rather than listwise data and can produce

⁵ Becker (2000) suggests using Fisher Z-values instead of correlations, particularly for primary samples that are small (n < 100). I used correlations based on justifications (p. 82-83) included in the Hunter and

biased or even inestimable results (cf. Darlington, 1990; Kline, 1998; Wothke, 1993). Because the major purpose of this study is to compare relational patterns within the overall model rather than to test the notion that job performance is completely determined by cognitive ability, personality, biodata, and interviews, the statistical significance of an overall model is not tested.

Schmidt method (2004), and because the sample size was usually fairly large.

RESULTS

Database Description

From 172 published studies, 195 conceptually "independent" samples and 984 unique correlations were obtained. The full listing of studies is in Appendix F. Table 3 provides a general breakdown of studies by type. Sample sizes ranged from N = 29 to N = 25,327 (for Hough's 1992 meta-analysis contributing a conscientiousness-task performance correlation), with an average of 812 subjects. Thus, the average sample size per sample was fairly large. Although there were about 5 unique correlations associated with each sample on average (after aggregating redundant measures and forming linear composites of measured subfacets), about half of the samples contributed a single correlation to the database. For those samples contributing more than one effect size, the average number of usable correlations reported per study is approximately 9. The majority of samples consisted of employees in nonmanagerial jobs and provided data about real job performance, versus simulated job tasks. Only 15% of the samples included participants who were not explicitly selected on one of the study variables, implying little to no range restriction for those samples.

The samples included in the database covered a wide range of jobs, and also included the general public (in some longitudinal studies) and university students. Because the criteria for study inclusion were broad, there are almost as many different job types as there are samples. Some of the job types included are manufacturing line workers, university administrative staff, hotel staff, agricultural coop employees, working students enrolled in an MBA program, telemarketers, food service workers, stockpersons,

account managers, pulp mill workers, expatriates in a technical company, computer programmers, summer camp workers, pharmaceutical workers, prison guards, and more. Some of the more commonly studied jobs were military soldiers, sales representatives, and insurance agents, often because the same researchers had access to the same organizations.

Regarding data needed to apply corrections for statistical artifacts, only one study required a correction for variable dichotomization. An "unsuitable discharge" variable representing "a failure to meet minimum behavioral or performance criteria" (McDaniel, 1989, p. 965) was included as task performance. In that sample, 16.5% of the employees were discharged for this reason.

Regarding corrections for measurement unreliability, 508 (73%) of 696 possible reliability estimates were obtained either from primary studies, test manuals, or other literature reporting statistics on the same measures. The type of reliability estimate differed for each measure in each study but the alpha coefficient was most commonly reported, followed by interrater reliability. Of the 185 missing values, 53 pertained to samples drawn from published meta-analyses that did not provide specific reliability information about each study. The actual number of reliability estimates obtained from the literature and number of imputed estimates are listed in Table 4.

The imputation process described earlier used the following reliability estimates of performance measures, based on work by Viswesvaran et al. (1996): task (.57), citizenship (.55), job dedication (.55), interpersonal facilitation (.47), overall (.81). These estimates are considerably lower than the mean values obtained in this study. Yet, a supplemental analysis using the mean values in database (Table 4) instead of the above

Database descriptives

Number of independent samples in database	195
Largest sample size	25,327
Smallest sample size	29
Average sample size	812
Total number of unique correlations in database	984
Average number of correlations contributed per sample	5
Number of samples providing only one correlation	97
Managerial	11 %
Nonmanagerial	65 %
Indeterminable or includes managers and lower	24 %
Sampled in a work setting	80 %
Range restricted	85 %

Note. The figures for "applied setting" and "range restricted" were based on the author's codes, as described in the methods section.

values produced nearly identical results. The specific number of estimates used in each analysis described below varied, but 23 of the imputed values that were associated with 48 interview correlations (column 4 of Table 4) were never used because too few studies were available in the literature (see below for further explanation).

Outlier Analyses

Before proceeding with meta-analytic computations, I checked that the data did not contain any obvious transcription errors (i.e., correlations above 1) and calculated the SAMD values for the observed correlations (Huffcutt & Arthur, 1995) to identify possible outliers. A scree plot of the absolute SAMD values was created for each correlation with at least 20 datapoints (Appendix G). Outliers are related to a number of issues including coding errors, model misspecification, and true score versus error. The plots shown here are simply meant to illustrate the distribution of recorded findings rather than to detect cases for removal since 1) there are not many cases per cell, 2) the point of meta-analysis is to determine if aberrant findings can be attributed to sampling error, and 3) the SAMD does not indicate the joint effect of multiple outliers within a study. That is, one outlier within a sample that otherwise provides correlations of "good quality" (i.e., non-outliers) is more likely be a true score than an outlier among many other outliers provided by the same sample, unless the lone error is due to a transcription mistake by the original authors or the meta-analytic coders, or due to a peculiar influence (e.g., an unreliable scale for that one measure).

Most of the plots show a desirable pattern with a plateau at the tail. Although a few plots showed some drops in the middle (e.g., conscientiousness – citizenship and openness – emotional stability), there was never a single point by itself after the initial

Reliability	Information	for Scales.

	Mean ·			Reported	Imputed	W/o
	Reliability	Min.	Max.	Estimates	Estimates	Interviews
Cognitive	.87	.46	.98	40	27	13
Extraversion	.81	.49	.94	37	6	5
Conscientious	.81	.62	.98	52	19	18
Agreeable	.76	.63	.97	41	8	7
Openness	.77	.55	.91	32	7	6
Emot. Stability	.83	.70	.92	35	8	7
Biodata	.79	.59	.91	15	7	7
Interview	.82	.59	.97	7	23	0
Task Perf	.77	.27	1*	59	23	21
Citizenship Perf	74	.32	.97	50	17	17
J. Dedication Perf	.83	.29	.99	61	12	12
Interpersonal Perf	.80	.31	.97	67	12	12
Overall Perf	.84	.42	.96	12	16	14
(Total)				(508)	(185)	(162)

Note. Reported Estimates = number of reliability estimates obtained from literature, Min. = minimum reliability estimate reported, Max. = maximum reliability estimate reported, Imputed Estimates = number of estimates imputed, W/o Interviews = number of estimates imputed, excluding interviews.

*Although the data are likely to contain some error, some objective criteria such as number of sales were assumed to have a reliability of 1, to ensure a conservative correction.

drop. Therefore, no explainable outliers were found and no cases were removed from the analyses.

Overview of Meta-analytic Results

The number of samples (k) and total sample sizes (N) for each relationship are presented above the diagonal in Table 5, where each cell represents an individual metaanalysis. The correlations above the diagonal are mean-weighted observed correlations uncorrected for statistical artifacts. Again, 95% confidence intervals were applied to uncorrected mean correlations. Bolded values have an interval that does not include 0, meaning that the effect is statistically significant. The correlations⁶ below the diagonal are means of correlations that have been corrected for the two statistical artifacts mentioned earlier. Also presented below the diagonal are 80% credibility intervals indicating the estimated true range of corrected correlations.

A single correlation was found for five cells: biodata-job dedication, biodatainterpersonal, interview-job dedication, citizenship-job dedication, and citizenshipinterpersonal. With the exception of its relationships to cognitive ability and task performance, only 2 studies with small samples contributed to the meta-analytic estimates for interviews. As a result, interviews were excluded from the remainder of the analyses since this meta-analysis would not be able to provide conclusions beyond those made in past primary studies and meta-analyses of employment interviews.

There were few studies providing information about biodata as well but enough to provide at least preliminary meta-analytic findings. Two of these studies (Hough et al.,

⁶ The mean corrected correlation (r_c) is often labeled ρ . I stray from convention because the label r_c is more informative, indicating how the estimate was derived, and because I attempt to derive "better" estimates of ρ with a multivariate meta-analysis. The multivariate method is conceptually superior but did not necessarily produce more accurate estimates because of practical limitations in the data (see Discussion).

1990; McHenry et al., 1990) used the Assessment of Background and Life Experiences (ABLE). The developers of the ABLE (Hough et al., 1990) intended to measure "temperaments." So, the ABLE might be viewed as a personality measure. However, it was classified as biodata, per this study definitions, because of the way it measures individual differences through past experience rather than preferences or intentions (as other personality tests can do).

Cognitive ability was weakly related to personality overall, with the highest mean corrected correlations being .19 with openness and .16 with emotional stability. In contrast, intercorrelations between the personality variables were higher than expected (e.g., compared to Hough, 1992) with the average r_c weighted by sample size equal to .38. In this dataset, emotional stability demonstrated the strongest links to other personality variables.

There were few studies measuring overall job performance due to the criteria for inclusion (requiring at least one dimensional measure of performance) and my research aims. The data that were collected failed to show the typically strong relationship between cognitive ability and overall performance. Even so, this estimate was based on 10 studies with a decent size (N = 8,009). These results do not contradict the large body of literature on the validity of general cognitive ability. They simply suggest that studies measuring overall performance in addition to specific performance dimensions will, for some reason, find lower validities.

On the other hand, the results resembled past findings of personality validities (e.g., Barrick et al. 2001; Salgado, 1998), but were generally of smaller magnitudes. Conscientiousness produced the largest mean corrected correlation (.20). In this set of

studies, overall performance was related to all performance dimensions as expected but slightly more strongly (p < .01) with citizenship ($r_c = .65$) than with task performance ($r_c = .41$).

More detailed statistics for uncorrected and corrected correlations that are of particular interest in this study are included in Table 6. Most of the mean confidence intervals excluded zero (typically for mean values above .06). Simultaneously, most of the credibility intervals for the mean corrected correlations were quite large, justifying a search for moderators.

Sampling error explained at least 60% of the observed variance in corrected correlations for just 7 relationships (Table 6). (The reason why some of the estimated values of V_{art} in Table 6 exceed 100 is mostly likely due to second order sampling error since most of these estimates involve a small number of studies.) Even so, complete homogeneity across the database correlations was not expected given the different types of samples included, the broadly defined constructs, and the small sample sizes (*k*) for some pairwise analyses that lowered the statistical power of moderator tests.

Based on the 60% rule, openness demonstrated the most stable estimates with various measures of performance but the effect sizes were essentially null. The mean corrected correlations were all below .10 except with overall performance. Emotional stability also produced small but homogenous correlations with job dedication ($r_c = .06$) and overall performance ($r_c = .07$), replicating meta-analytic findings in Hurtz and Donovan (2000). The correlation between extraversion and task performance was also small (.03) but homogeneous. The Q values and their p values generally supported the same conclusions as those supported by the size of the credibility intervals.

Meta-analytic C	Correlation Matrix for	or Job Performance	and Performance Predictors.

		1	2	3	4	5
(1)	Cognitive		0 (07, .07)	.05 (.01, .09)	0 (05,.06)	.16 (.10, .23)
			14, 4571	28, 11208	14, 6686	14, 5029
(2)	Extraversion	.01		.22 (.16, .28)	.24 (.17, .31)	.26 (.20, .33)
		(16, .18)		34, 9624	30, 8943	28, 8356
(3)	Conscientious	.06	.28		.38 (.32, .44)	.10 (.03, .17)
		(09, .21)	(0, .56)		38, 17574	30, 9363
(4)	Agreeable	0	.30	.49		.15 (.09, .22)
		(14, .15)	(04, .64)	(.17, .81)		28, 8356
(5)	Openness	.18	.33	.12	.19	
		(0, .36)	(.07, .58)	(21, .45)	(11, .50)	
(6)	E. Stability	.16	.38	.62	.50	.17
		(06, .38)	(.06, .69)	(.27, .96)	(.14, .87)	(02, .37)
(7)	Biodata	.27	.31	.37	.37	.44
		(.06, .48)	(.18, .44)	(.29, .45)	(.30, .44)	(.13, .75)
(8)	Interview	-	-	-	-	-
(9)	Task	.28	.03	.09	.04	.02
(-)	2007	(0, .56)	(05, .12)	(.01, .18)	(04, .13)	(05, .08)
(10)	Citizenship	.29	.06	.20	.16	.06
()	•••••• <i>P</i>	(.06, .51)	(11, .22)	(.09, .32)	(.04, .28)	(01, .13)
(11)	Dedication	.09	0	.17	.15	.06
()		(0, .18)	(14, .14)	(.03, .31)	(.04, .26)	(.06, .06)
(12)	Interpersonal	.04	.03	.13	.20	.01
(12)	incipei sonui	.04 (06, .14)	.05 (05, .12)	(.02, .24)	(.10, .31)	(.01, .01)
(13)	Overall	.11	.01	.20	.06	.11
(13)	Gveruii					
		(.04, .18)	(15, .16)	(.09, .31)	(05, .16)	(.11, .11)

Note. Italicized variables are job performance. Information above the diagonal includes the mean weighted correlation (r), the 95% confidence interval in parentheses, the number of studies (k), and the total sample size for that estimate (N) across samples. Information below the diagonal includes the mean weighted correlation corrected for artifacts (r_c) , and the 80% credibility interval in parentheses. Bolded confidence intervals exclude 0.

Table 5 (continued)

		6	7	8	9	10
(1)	Cognitive	.14 (.07, .22)	.23 (.14, .32)	.20 (.14, .27)	.19 (.13, .25)	.22 (.15, .28)
		15, 8226	9, 16610	18, 6048	30, 42107	15, 17430
(2)	Extraversion	.31 (.24, .39)	.25 (.16, .34)	.18 (.17, .19)	.03 (02, .08)	.05 (0, .11)
		29, 8774	4, 1010	2, 1148	14, 2651	20, 4425
(3)	Conscientious	.50 (.43, .56)	.27 (.24, .31)	.18 (.15, .21)	.08 (.06, .10)	.15 (.12, .18)
		34, 17821	8, 6429	2, 3625	21, 38787	25, 15425
(4)	Agreeable	.40 (.32, .47)	.28 (.24, .32)	.20 (0, .41)	.03 (01, .07)	.12 (.08, .16)
		31, 16110	6, 5978	2, 437	13, 3919	20, 6143
(5)	Openness	.14 (.09, .19)	.34 (.13, .56)	.13 (.05, .22)	.02 (04, .08)	.05 (0, .09)
		28, 9563	4, 859	2, 620	10, 1367	15, 2930
(6)	E. Stability		.35 (.30, .40)	.24 (.16, .32)	.08 (.05, .11)	.08 (.05, .11)
			8, 6697	2, 1010	11, 10323	15, 11782
(7)	Biodata	.45		.16 (.02, .30)	.12 (.09, .15)	.21 (.16, .26)
		(.34, .57)		2, 1038	15, 44904	10, 14200
(8)	Interview	-	-		.16 (.11, .21)	.28 (.18, .39)
					6, 7493	4, 827
(9)	Task	.11	.17	-		.42 (.34, .5)
		(.03, .19)	(.05, .29)			48, 22276
(10)	Citizenship	.10	.30	-	.49	
		(.01, .20)	(.16, .44)		(.04, .95)	
(11)	Dedication	.06	-	-	.39	-
		(02, .15)			(06, .84)	
(12)	Interpersonal	.06	-	-	.35	-
		(04, .17)			(07, .78)	
(13)	Overall	.07	.26	-	.41	.65
-		(0, .15)	(.13, .39)	_	(.15, .67)	(.35, .94)

Tabl	le 5 ((continued)	
------	--------	-------------	--

-		11	12	13
(1)	Cognitive	.08 (.02, .13)	.04 (01, .09)	.09 (.06, .13)
		5, 2501	6, 3118	10, 8009
(2)	Extraversion	.01 (07, .08)	.03 (03, .08)	0 (08, .08)
		9, 2026	10, 2534	9, 1941
(3)	Conscientious	.12 (.07, .18)	.10 (.05, .14)	.16 (.10, .23)
		12, 4272	12, 4713	9, 1941
(4)	Agreeable	.11 (.06, .16)	.15 (.10, .19)	.04 (02, .11)
		11, 4205	12, 4713	8, 1584
(5)	Openness	.05 (01, .11)	.01 (03, .04)	.09 (.04, .14)
		6, 967	8, 1742	7, 1076
(6)	E. Stability	.05 (02, .13)	.06 (01, .12)	.06 (0, .12)
		5, 924	7, 1699	7, 1390
(7)	Biodata	.25 (n/a)	.30 (n/a)	.20 (.11, .28)
		1, 116	1, 368	4, 6020
(8)	Interview	.36 (n/a)	.18 (.11, .25)	.22 (.17, .26)
		1, 47	3, 366	3, 349
(9)	Task	.36 (.26, .46)	.32 (.24, .40)	.29 (.20, .39)
		24, 8168	28, 9720	14, 9701
(10)	Citizenship			.54 (.39, .69)
		-	-	10, 3547
(11)	Dedication		.60 (.56, .65)	.55 (.45, .65)
			57, 17360	11, 3432
(12)	Interpersonal	.72		.48 (.32, .63)
		(.47, .96)		10, 3316
(13)	Overall	.65	.59	
		(.42, .87)	(.28, .91)	

Finally, estimates of the true population correlation, ρ , were calculated using corrected correlations in each study according to the multivariate procedure described in the Method section. I estimated the matrix represented by the model in Figure 1 using 661 correlations taken from 115 independent samples. As stated earlier, I needed the full set of intercorrelations between all variables (relevant to this meta-analysis) studied within each sample, though studies could contribute different numbers and types of correlations to the analysis. Based on the 20% cutoff rule I chose, it was necessary to impute one value using the mean value of that correlation in the total database.

The main strength of this multivariate approach to meta-analysis, at least in theory, is that comparisons of relational patterns will be more accurate because they are weighted by the variance-covariance matrix representing dependencies among correlations within each sample. A caveat to these estimates being more accurate is that the studies contributing data to the final estimates must still be representative of the true population of relevant studies. When certain types of studies are excluded from or are overrepresented in the analysis, there is always the potential for introducing bias and creating model misspecification (Raudenbush et al., 1988).

Figure 4 shows the estimates of ρ produced by weighting corrected correlations by sample variance-covariance matrices. The paths depicted generally were similar to the results produced by separate meta-analyses in Table 6. However, biodata validities were considerably higher and extraversion became a weak but noticeable predictor of citizenship. Also, the magnitude of emotional stability validities rose slightly while cognitive ability became a weaker predictor of citizenship performance. Finally, some of the intercorrelations between predictors were unpredictably high (Table 7) which

Meta-analytic Results For Pairs of Study Variables.

Study	k	r	95% CI	r _c	SDrc	80% CV	V _{art} (%)	Q
Performances				<u>v</u>				£
Task – CitizenP	48	.42	.34, .50	.49	.36	.04, .95	2	24332.01
Task – DedicatP	24	.36	.26, .46	.39	.35	06, .84	3	9012.43
Task – InterP	28	.32	.24, .40	.35	.33	07, .78	4	9 765.78
Dedicat – InterP	57	.60	.56, .65	.72	.19	.47, .96	5	8965.43
Overall Performance	e							
Task	14	.29	.20, .39	.41	.20	.15, .67	5	1523.21
Citizenship	10	.54	.39, .69	.65	.23	.35, .94	4	788.41
Job Dedication	11	.55	.45, .65	.65	.18	.42, .87	6	878.89
Interpersonal	10	.48	.32, .63	.59	.25	.28, .91	4	503.19
Task Performance								
Cognitive	30	.19	.14, .25	.28	.22	0, .56	3	2679.13
Extraversion	14	.03	02, .08	.03	.07	05, .12	62	31.83
Conscientiousness	21	.08	.06, .10	.09	.07	.01, .18	15	255.86
Agreeableness	13	.03	01, .07	.04	.07	04, .13	56	29.45
Openness Emotional	10	.02	04, .08	.02	.05	05, .08	79	18.20*
Stability	11	.08	.05, .11	.11	.06	.03, .19	34	41.57
Biodata	15	.12	.09, .15	.17	.10	.05, .29	7	462.19
Citizenship Perform	ance							
Cognitive	15	.22	.15, .28	.29	.18	.06, .51	4	604.64
Extraversion	20	.05	0, .11	.05	.12	11, .22	30	107.84
Conscientiousness	25	.15	.12, .18	.20	.08	.09, .32	3	128.02
Agreeableness	20	.12	.08, .16	.16	.10	.04, .28	36	82.02
Openness Emotional	15	.05	0, .09	.06	.06	01, .13	72	34.04
Stability	15	.08	.05, .11	.10	.07	.01, .20	4	74.33
Biodata	10	.21	.16, .26	.30	.11	.16, .44	10	223.36

Note. k = # of samples; r = uncorrected weighted average correlation; 95% CI = confidence interval around r; $r_c =$ corrected weighted average correlation; $SDr_c =$ standard deviation of r_c ; 80% CV = credibility interval around r_c ; $V_{art} =$ percentage of r_c variance explained by study artifacts; Q = homogeneity statistic. Bolded V_{art} supports homogeneity using the 60% rule. *p <.01; **p < .05. Dashes are values not estimated.

Study	k	r	95% CI	r	SDr _c	80% CV	V _{art} (%)	Q
	<u> </u>	/		r _c	JUIC	00/00 4	/ツ	¥
Job Dedication Perfo	rmanc	<u>e</u>						
Cognitive	5	.08	.02, .13	.09	.07	0, .18	40	17.06
Extraversion	9	.01	07, .08	0	.11	14, .14	35	44.94
Conscientiousness	12	.12	.07, .18	.17	.11	.03, .31	28	84.94
Agreeableness	11	.11	.06, .16	.15	.09	.04, .26	39	45.33
Openness	6	.05	01, .11	.06	.00	.06, .06	117	9.50**
Emotional Stability	5	.05	03, .13	.06	.07	02, .15	67	12.90*
Biodata	1	.25	-	.31	.00	-	-	-
Interpersonal Perform	nance							
Cognitive	6	.04	01, .09	.04	.08	06, .14	37	21.98
Extraversion	10	.03	03, .08	.03	.07	05, .12	56	27.93
Conscientiousness	12	.10	.05, .15	.13	.09	.02, .24	38	46.28
Agreeableness	12	.15	.10, .19	.20	.08	.10, .31	40	47.64
Openness	8	.01	03, .04	.01	.00	.01, .01	144	9.10**
Emotional Stability	7	.06	01, .12	.06	.08	04, .17	48	23.42
Biodata	1	.30	-	.48	.00	-	-	-
Overall Performance								
Cognitive	10	.09	.06, .13	.11	.05	.04, .18	36	41.52
Extraversion	9	0	08, .08	.01	.12	15, .16	32	41.81
Conscientiousness	9	.16	.10, .23	.2	.09	.09, .31	47	3.74
Agreeableness	8	.04	02, .11	.06	.08	05, .16	53	25.59
Openness	7	.06	.00, .12	.07	.06	0, .15	70	15.12*
Emotional Stability	7	.09	.04, .14	.11	.00	.11, .11	153	7.78**
Biodata	4	.20	.11, .28	.26	.10	.13, .39	9	8.80
Cognitive Ability								
Extraversion	14	0	06, .06	.01	.14	16, .18	20	119.04
Conscientiousness	28	.05	.01, .09	.06	.12	09, .21	21	208.76
Agreeableness	14	.01	05, .06	0	.11	14, .15	2	112.96
Openness	14	.15	.09, .21	.18	.14	.00, .36	17	134.47
Emotional Stability	15	.13	.05, .21	.16	.18	06, .38	8	278.68
Biodata	9	.23	.14, .32	.27	.16	.06, .48	3	507.52

Table 6 (continued)

Table 6 (continued)

Study	k	r	95% CI	r	SDr _c	80% CV	V _{art} (%)	0
Other Predictors			3370 CI		5DIc	00/001		<u>¥</u>
Extrav – Conscient	34	.22	.16, .28	.28	.22	.00, .56	10	2874.98
Extrav – Agreeable	30	.24	.17, .31	.30	.26	04, .64	7	3946.74
Extrav – Openness	28	.26	.20, .33	.33	.20	.07, .58	11	2916.71
Extrav - Emot Stab	29	.31	.24, .39	.38	.25	.06, .69	6	1686.15
Extrav – Biodata	4	.25	.16, .34	.31	.10	.18, .44	36	19.09
Consc – Agreeable	38	.38	.32, .44	.49	.25	.17, .81	4	3521.78
Consc – Openness	30	.10	.03, .17	.12	.26	21, .45	7	1585.42
Consc - Emot Stab	34	.50	.43, .56	.62	.27	.27, .96	2	9525.02
Consc – Biodata	8	.27	.24, .31	.37	.06	.29, .45	34	45.20
Agree – Openness	28	.15	.09, .22	.19	.24	11, .50	9	771.51
Agree - Emot Stab	31	.40	.32, .47	.50	.28	.14, .87	3	4356.39
Agree – Biodata	6	.28	.24, .32	.37	.05	.30, .44	34	28.24
Open - Emot Stab	28	.14	.09, .19	.17	.15	02, .37	16	408.05
Open – Biodata	4	.34	.13, .56	.44	.24	.13, .75	9	97.13
Emot St – Biodata	8	.35	.30, .40	.45	.09	.34, .57	16	91.14

warrants some caution in interpreting these results, with eight larger than .95. See the supplemental analyses below for further explanation.

In conclusion, cognitive ability, conscientiousness and biodata were the best predictors of the two performance dimensions based on both the pairwise and multivariate meta-analyses. All predictors were related to biodata. The specific study hypotheses are evaluated in the following sections based on these results and some additional analyses.

Hypothesis 1: Moderation by Job Type

Hypothesis 1 (H1) predicts that the correlation between citizenship performance and noncognitive predictors will vary depending on whether jobs are managerial and/or sales related versus other types of jobs. Sample level codes for managerial (vs. lower) and sales (vs. other) jobs were assigned during the initial coding phase. Of 152 samples to receive managerial codes, 26 consisted primarily or completely of managerial jobs based on the information provided in the primary studies. Of the 140 samples to receive sales codes, 19 consisted primarily or completely of sales jobs.

H1 was based on the assumption that citizenship behaviors are a central part of both managerial and sales jobs, and that this common confound causes the moderation. Thus, H1 was tested using a single dichotomous category distinguishing the 45 managerial and/or sales job samples from the 99 other job samples. The first moderator analysis included the Big Five as the "noncognitive" predictors, while the second included the Big Five and biodata.

Fifty six correlations involving the Big Five and citizenship were available in 16 samples. Seven of these were managerial / sales jobs. Because multiple correlations

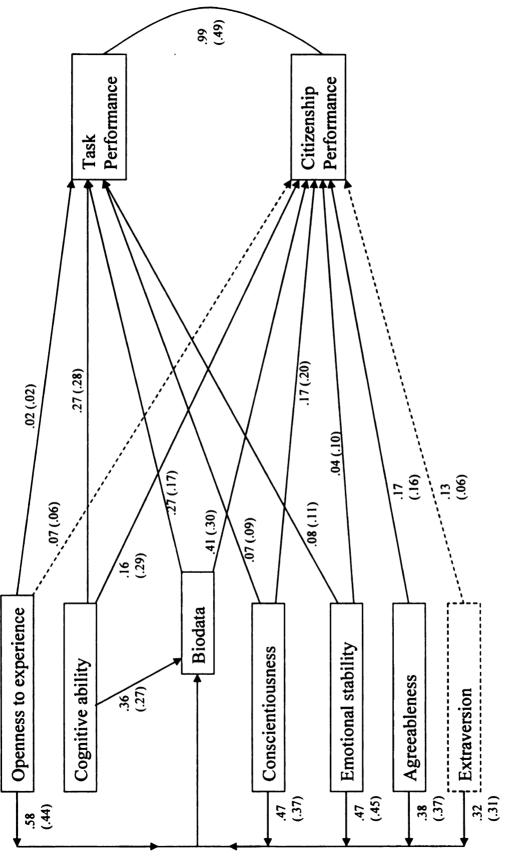
provided by a study for each personality dimension are dependent, I aggregated results within studies either by forming a linear composite correlation when interrcorrelations between personality dimensions were available or by averaging correlations (which is the lower bound of the linear composite).

The results of this analysis are shown in Table 8. The mean corrected correlation for that was different between the two groups but, contrary to H1, was weaker for managerial / sales jobs (.16 versus .31 for other jobs). The SD_{rc} for the combined sample was reduced in both subgroups, meaning that the credibility intervals became smaller. The between-groups Q_b of 57.88 was statistically significant at p < .01. Together the results of these moderator tests support managerial / sales job types as a moderator of the personality-citizenship relationship but in the opposite direction of that hypothesized. Also, the percentage of variance explained by artifacts was 80% for managerial / sales jobs but did not change much for "other" job types.

The results of the second analysis after adding biodata as a noncognitive predictor are similar (Table 8) except that a smaller percentage of the variance in mean corrected correlations was attributable to statistical artifacts in both groups as well as the total group. This is not surprising since biodata typically differ from personality measures and may assess some part of cognitive ability to a greater degree. In conclusion, H1 was not supported but there was evidence to support managerial / sales job type as a moderator.

It is noted that these results could be biased because the set of correlations reported for one subgroup were not necessarily related to the same personality dimensions as the correlations reported for the other subgroup; aggregating correlations across personality dimensions could mask a confound between job type and personality

Figure 4. Path Diagram with Multivariate Estimates of Study Variables.



Note. Estimates of individual meta-analyses (Table 6) are in parentheses. Dashed boxes and lines indicate variables and paths not included in Figure 1.

sis.	
aly	
a-ar	
Met	
ate l	
varia	
ultiv	
M	
o p	
ase	
ix E	
Aatr	
on N	
latio	
orre	
nC	
latio	
opul	
d P	
nate	
Stin	
щ	

		1	7	m	4	S	9	7	×
	Cognitive Ability								
ଲ	Extraversion	.216							
$\widehat{}$	Conscientiousness	.121	1.000						
Ŧ	Agreeableness	.075	1.000	989.					
5	Openness	.296	003	.952	.396				
9	Emotional Stability	.177	1.000	.984	.981	.304			
E	Biodata	.358	.319	.474	.376	.581	.470		
8	Task Performance	.275	.054	.069	.053	.018	.080	.266	
6	Citizenship Perf.	.158	.123	.169	.171	.071	.039	.409	988.

dimension. (Ideally, all studies would have provided correlations between all relevant variables and this confound would be controlled.) If some personality dimensions tend to produce correlations of very different magnitude from other dimensions *and* studies of one job type tend to report correlations for a particular set of personality dimensions, the results would incorrectly support a moderator. As a hypothetical example, personality studies of managers might measure extraversion more often than in studies of automotive line workers, which themselves tend to measure conscientiousness more often. If conscientiousness is a better predictor of citizenship than extraversion regardless of job type, effect sizes would show variance supporting moderation but not for the suspected reason of job type.

A closer examination, however, suggests that job type does moderate relationships here. Table 9 shows the number of correlations contributed to the subgroups by each personality dimension. The data for the managerial / sales group were composed of a greater number of conscientiousness and agreeableness correlations, and these correlations were higher than those for the other personality variables. The opposite compositional pattern is true for the non-managerial / sales group. Therefore, one would expect managerial / sales jobs to show stronger personality-performance correlations under the assumption that job type was confounded with the dimensions of personality measured. I found the opposite pattern of results (Table 8), supporting job type as a moderator, rather than the type of personality dimension measured.

Hypothesis 2: Moderation for Citizenship Dimension

Hypothesis 2 (H2) predicts that distinguishing interpersonal facilitation from job dedication for measures of citizenship will produce two unique patterns of relationships with other variables. One approach for evaluating this would be to group correlations between citizenship and other variables based on some estimate of how much the citizenship measure captures either subdimension. In essence, that is what the correlations for job dedication and interpersonal facilitation in Tables 5 and 6 represent. During the coding process, raters attempted to categorize performance correlations into interpersonal facilitation or job dedication. Correlations were categorized into the broader variable of citizenship when not clearly assessing one dimension or the other. This coding process effectively created four categories to which a sample could be assigned based on the citizenship correlations reported: job dedication, interpersonal facilitation, both job dedication and interpersonal facilitation, and overall citizenship assumed to measure job dedication and interpersonal facilitation to some unknown degree.

A traditional moderator analysis comparing the four types of samples could not be conducted due to small subgroup sizes. If mutually independent categories were created, there would be just four studies for job dedication and nine for interpersonal facilitation. Twenty six studies reported data for both citizenship dimensions separately but would have to be aggregated to preserve independence between effect sizes, resulting in the loss of crucial information. Consequently, I decided to evaluate H2 in a more qualitative fashion by examining meta-analyses using citizenship performance (in Table 6) as compared with the results of meta-analyses using job dedication and interpersonal facilitation separately, as has been done in past research (e.g., Hurtz & Donovan, 2000).

Because the majority of studies examined provided correlations for both subdimensions of citizenship, I treated correlations as if they were obtained from a single sample (using the smaller N between relevant cells in Table 5) and conducted (two-tailed)

Tests of	<u>t</u> t	ıe	Moderating	Effect	of	Job	Type.
	_						the second s

Grouping	<u>k</u>	N	r _c	SD_{rc}	Q	% V _{art}
			D			
			Persona	lity-Citiz	enship	
Managerial / Sales	7	2924	.16	.03	11.24	80.2
Other	9	10466	.31	.06	58.79	17.9
Combined sample	16	13390	.29	.08	127.91	15.6
$\mathcal{Q}_{ extsf{b}}$					57.88*	
		Perso	nality &	Biodata	Citizenshi	p
Managerial / Sales	8	3501	.14	.05	19.87	52.9
Other	14	23079	.26	.08	184.14	7.9
Combined sample	22	26580	.25	.09	248.24	9.7
\mathcal{Q}_{b}					44.26*	

Note. k = number of samples used in analysis, N = total sample size for analysis, r_c = mean corrected correlation, SD_{rc} = standard deviation of r_c , Q = homogeneity statistic, Q_b = difference total Q for combined sample and sum of Qs for subgroups, % V_{art} = percentage of variance due to statistical artifacts. *p < .01

Table 9

Percentage of Correlations From Each Personality Dimension

	Manager / Sales	Other
Extraversion	16	22
Conscientiousness	37	24
Agreeableness	26	22
Openness	11	16
Emotional Stability	11	16

t-tests of the difference in correlations by dimension, for each predictor and for task performance. Table 10 provides the results of these tests. There was a significant difference for five of the seven variables at p < .05. In looking at the actual correlations, the magnitude of the difference in correlations was quite small in most cases. Given that these estimates were corrected for measurement unreliability in the both variables, there may not be a practically meaningful difference in validities across the two types of citizenship dimensions.

In addition, I examined the change in the credibility interval size for overall citizenship and the subdimensions. The interval shrank by more than 50% for cognitive ability when subdimensions were used. For openness, the intervals shrank to essentially 0 when analyzed by subdimension. The rest of the correlations had intervals that were consistently large irregardless of performance measure.

Another very important piece of information to consider is the correlation between the citizenship dimensions. The estimated population correlation for the two dimensions is very high ($r_c = .72$) but not to the point of complete overlap, especially considering that this estimate was corrected for some artifactual variance. Also, this estimate is probably biased upwards due to common method variance since both measures were often subscales of the same instrument. Nonetheless, a strong relationship was expected since the dimensions are both indicators of citizenship in theory. The important issue is whether the two constructs are so strongly related that they are functionally redundant.

To conclude H2 could not be tested directly but the pattern of results from individual tests suggests that even if there is some statistical support for H2, the

Simple Com	parisons o	of Job Dec	dication and	Interpersonal	Facilitation	Effect Sizes.

	Job Ded.	Interpersonal			
Correlate	$(r_{\rm c})$	$(r_{\rm c})$	N	t	р
Cognitive	.09	.04	2501	3.43	0.001
Extraversion	0	.03	2026	1.84	0.066
Conscientiousness	.17	.13	4272	3.66	0.000
Agreeableness	.15	.20	4205	4.60	0.000
Openness	.06	.01	947	2.10	0.036
Emotional Stability	.06	.06	924	0.00	1.000
Task Performance	.39	.35	8168	5.87	0.000

Note. N is the smaller sample size of the two groups.

distinction between job dedication and interpersonal facilitation is not practically meaningful. Examining effect sizes based on the proportion of interpersonal facilitation or job dedication measured does not appear to improve validity estimates substantially. These results are similar to those produced by earlier studies (i.e., Conway, 1999; Hurtz & Donovan, 2000).

Hypothesis 3 and 4: Differential Prediction Patterns for Performance Dimensions

Although no hypothesis was formed due to a lack of consistent findings in earlier studies, the correlation between task and citizenship performance is one of the most interesting findings and useful contributions of this study. Finding evidence that the two dimensions are distinct is a perquisite for testing H3 and H4, as well as H7 through H12 later. The two dimensions were moderately correlated but this estimate was unstable ($r_c = .49$; $SD_{rc} = .36$). Accordingly, past researchers have varied substantially in their estimates of this relationship. Nevertheless, this finding strongly suggests that the relationship is substantial for many cases and should not be ignored; it appears to be nonzero based on confidence and credibility intervals.

From the approach of understanding a construct with a nomological network, evaluations of H3 and H4 should shed more light on this issue. H3 specifically predicts that task performance will be more strongly related to cognitive ability than personality. The data associated with cognitive ability and each personality predictor in Tables 5 and 6 are based on different subsets of primary studies in the database. Thus, the multivariate estimates shown in Figure 4 and Table 7 provide the best evaluation of this prediction (refer to Method section about testing linear models), although the results from separate meta-analyses (Table 6) are expected to show some general convergence. Cognitive

ability is the strongest predictor of task performance ($\rho = .27, r_c = .28$)⁷ compared to any other single personality dimension. Conscientiousness and emotional stability produced statistically significant (p < .05) but weak relationships. The mean effect size across personality dimensions is less than .10, using either ρ or r_c estimates. There is no direct way to test the significance of the difference in relationships given the various sample sizes used in calculations but the difference is clearly substantial and, presumably, not caused by sampling error since the total N for the separate meta-analyses was typically very large.

H4 predicts that citizenship will be more strongly related to personality than cognitive ability. The results pertaining to individual personality dimensions varied somewhat depending on whether pairwise or multivariate estimates are considered. Extraversion and openness produced the lowest mean corrected correlations (both = .06) but emotional stability and openness produced the lowest ρ 's after controlling within study dependencies. Nonetheless, cognitive ability was the best predictor of citizenship performance in both types of analyses, although ρ (but not r_c) for conscientiousness is equally strong. The effect size of citizenship with overall personality (treated as a class of measures but not as a unitary construct) is about .12 based on either the mean ρ or r_c across personality dimensions. Cognitive ability, on the other hand, produced a larger correlation with citizenship ($\rho = .16$, $r_c = .29$), thereby failing to support H4.

Hypotheses 5 and 6: Moderation by Job Complexity

Hypotheses 5 and 6 concern the possible moderation of correlations between cognitive ability and dimensions of job performance by job complexity. I had originally

⁷ Again, these are estimates of the same relationship but r_c is more commonly reported while " ρ " is more theoretically sound.

intended to use codes indexing job complexity obtained from the O*NET database. The database did not provide the necessary data to a sufficient degree. O*NET lacked codes for many of the jobs studied, including military jobs and nonspecific jobs classified as "residual jobs." In all, just 61 codes were obtained. Moreover, low level jobs all received a code of "Below 4.0" and were not distinguishable from each other. As a result, I created a dichotomous moderator indexing high vs. low job complexity by splitting the obtained O*NET SVP scores. Scores below 6 were recoded as low complexity while scores 6 or higher became high complexity. Of the 33 studies providing the relevant correlations for testing H5 and H6 (i.e., cognitive ability – task performance and cognitive ability – citizenship performance), 2 did not provide enough information about their samples and 7 did not have SVP codes to be recoded.

For the samples still missing codes, I assigned job complexity codes through a rating process. Seven graduate students in psychology, including the author, read a brief description of the samples. The raters were then asked to consider the total range of jobs encountered in research literature (e.g. toll booth operator to medical physician) and coded the current samples as either high (1), low (0), or indeterminate. These data are included in Appendix H. When treating "indeterminate" ratings as missing data, the interrater reliability (i.e., KR-20 coefficient) was .94 and the average kappa value between all rater pairs was .49. The average kappa is typically a good approximation of multi-rater agreement indices like Light's κ_m (Conger, 1980). Three studies showed 100% agreement across the raters. Two studies (Hedge & Teachout, 1992 and Ree, Earles & Teachout, 1994) each received two indeterminate ratings. For the former study, 4 of the 6 remaining ratings were low complexity. For the latter study, all of the

remaining 6 ratings were high complexity. Due to the relatively good agreement reached, the mean rating across raters was used as the job complexity code.

Contradictory to Hypotheses 5, none of the moderator tests (Table 11) supported the notion that the job complexity moderates the correlation between cognitive ability and task performance. It would be wrong, however, to accept this null hypothesis and conclude that job complexity is definitely not a moderator.

Contrary to H6, the results (Table 11) support job complexity as a moderator of cognitive ability -citizenship performance correlations. The credibility intervals were smaller for subgroups as indicated by the SD_{rc} values and Q_b was statistically significant. The difference in mean corrected correlations for the subgroups was fairly large, with cognitive ability being a good predictor of citizenship in less-complex jobs and a very weak predictor in complex jobs.

Hypotheses 7 through 10: Specific Predictor – Criterion Relationships

The next set of hypotheses predicted various types of relationships between the predictors and the job performance dimensions. The relevant statistics for this section can be found in Tables 5, 6, and 7. Tests of statistical significance are tied to whether 0 is included in 1) the 95% confidence interval for the uncorrected mean correlation, and 2) the 80% credibility interval for the corrected correlation (or for ρ). For all tests of Hypotheses 7 through 10, the conclusions implied by the confidence intervals agreed with those implied by the credibility intervals, and the interval test results corresponded to tests of each ρ in the multivariate analysis. (An exact test of ρ and its pooled standard error could not be computed with the current multivariate method because every study did not provide a correlation for every study variable.)

Table 11

Grouping	k	N	r _c	SD_{rc}	Q	% V _{art}
			Co	gnitive-T	<u>ask</u>	
Complex Jobs	8	7533	.24	.13	99.74	11.2
Less-complex Jobs	14	29431	.21	.27	2285.89	0.9
Combined sample	22	36964	.21	.25	2385.64	1.4
\mathcal{Q}_{b}					.26	
			Cognit	ive-Citiz	enship	
Complex Jobs	6	3069	.09	.07	20.19	40.4
Less-complex Jobs	6	13412	.37	.13	288.74	4.5
Combined sample	12	16481	.31	.16	512.91	4.5
\mathcal{Q}_{b}					203.98*	

Tests of the Moderating Effect of Job Complexity.

Note. k = number of samples used in analysis, N = total sample size for analysis, $r_c =$ mean corrected correlation, $SD_{rc} =$ standard deviation of r_c , Q =homogeneity statistic, $Q_b =$ difference total Q for combined sample and sum of Qs for subgroups, % $V_{art} =$ percentage of variance due to statistical artifacts. *p < .01

Although significance testing can aid interpretation, specific tests do not seem to be as meaningful for meta-analytic studies because point estimates are derived (under the fixed-effects model), after accounting for first-order sampling error. It seems more relevant to instead consider the magnitude of the correlation and whether it is practically significant. In the absence of more specific standards, one can always refer to Cohen's (1977) criteria: small (but still meaningful) effect size = at least .10, medium = at least .30, and large = .50 or higher.

Hypothesis 7 predicted that conscientiousness would be significantly related to task and citizenship performance. Conscientiousness was significantly related to task ($\rho = .07$, total N = 28,040) and citizenship ($\rho = .17$, total N = 3,250) performance, and the difference between the correlations was statistically significant at p < .01 using the z-test for Fisher-transformed values.⁸ These results support H7. It was not necessary to test Hypothesis 7B given the lack of support for a general distinction between citizenship dimensions. Nevertheless, for those interested specifically in conscientiousness and not personality generally, the results in Table 5 support the idea that conscientiousness acts through job dedication.

The data do not support Hypothesis 8, that emotional stability is related to both dimensions of performance, but more strongly with citizenship. Emotional stability correlated weakly with both task ($\rho = .08$) and citizenship ($\rho = .04$) performance. The difference between correlations was not statistically significant (p > .05) and the credibility intervals for both values overlapped almost completely. H8B was not formally evaluated but the statistics in Table 5 suggest that emotional stability is equally

⁸ The z-test is not perfectly appropriate because the samples are not completely independent. At the other extreme, a one-sample t-test using the smallest sample size also produced a significant difference at p < .01.

(statistically) unrelated to job dedication and interpersonal performance.

In contrast, strong support for was found Hypothesis 9 as the ρ for agreeableness was .17 with citizenship but just .05 with task performance. The difference between these correlations was also statistically significant (z = 3.98, p < .01). Hurtz and Donovan (2000) found similar results in that agreeableness showed higher correlations with interpersonal facilitation than with task performance. To examine the possibility that agreeableness influences citizenship specifically through interpersonal facilitation, I ran a supplemental multivariate meta-analysis to make valid comparisons. The results did not support the notion; the ρ 's for both citizenship facets with agreeableness were about .22.

Hypothesis 10 was not supported. Openness was not significantly related to task or citizenship performance.

In summary, conscientiousness, agreeableness, and extraversion were fair predictors of citizenship performance while none of the personality dimensions were good predictors of task performance.

Hypothesis 12: Prediction of Biodata Linked to Constructs

Hypothesis 12 predicts that degree to which biodata assess personality and cognitive ability will determine their validity with task and citizenship performance. To make comparisons between variables using different sets of studies, it was necessary to account for dependence between reported effect sizes since some studies reported both biodata-task and biodata-citizenship correlations while others reported one or the other. Consequently, I estimated the mean corrected correlations of biodata with task and citizenship performance, weighting correlations by each sample's variance-covariance matrix according to the multivariate method described previously. These data were then

analyzed in a generalized least squares regression model using the "proportion" (of constructs assessed) code to predict variation in the magnitude of correlations across studies (Raudenbush et al., 1988). The biodata "proportion" codes assigned to relevant samples in the third stage of coding (described earlier and included in Appendix I) refer to the proportion of an entire biodata measure that assesses cognitive ability when testing H12A, and to the proportion of biodata that assesses personality (excluding extraversion) when testing H12B. (To avoid losing information through the dichotomization of variables, these proportion values were used to test hypotheses rather than some arbitrary cutoff to determine whether biodata "primarily" assessed one construct or another.)

I then analyzed the 14 eligible studies (total N = 28,500) providing 29 correlations, after mean imputing three values for the correlation between task and citizenship. The analyses involved regressing task performance-biodata and citizenshipbiodata correlations on 1) the proportion of biodata assessing cognitive ability and 2) the proportion biodata assessing personality. Significance (*z*-) tests using standard errors from the pooled variance-covariance matrix were conducted to examine whether the regression slopes in this model were nonzero. (Therefore, there is no single *N* associated with the significance tests, just a pooled estimate of the variance for each effect size.) The results are as follows.

In predicting task performance with biodata, the regression results support H12A. The more that a biodata measures cognitive ability versus other factors, the stronger biodata validities become, by a factor of .014 for every "correlational unit" increase in the proportion of cognitive ability. This slope value of .014 is small but statistically significant (z = .014 / .00159 = 9.05, p < .01). Interestingly, ratings of proportion indicate that cognitive ability was never associated with more than 20% of a biodata measure.

In predicting citizenship performance, the results fail to support H12A. The proportion of biodata assessing personality had a statistically significant slope value of - .0003 (z = 5.04, p < .01) but the effect was too small to be of practical importance. Thus, H12B was not supported in any meaningful sense. Hypothesis 12C could be tested because there were only a few samples, none used biodata that even came close to measuring cognitive ability and personality in equal proportions. Despite not having produced a clear link between predictor constructs and job performance, the biodata validities estimated here are relatively high, particularly for citizenship ($\rho = .41$).

Supplemental Analyses

I ran some supplemental analyses to check the sensitivity of the various metaanalytic results. First, I investigated how the two largest studies affected overall estimates. These studies were not considered to be true outliers but were more influential in deriving a cumulated estimate. Hough (1992) had a total sample size of 25,327 to estimate the correlation between conscientiousness and task performance in her metaanalysis, and Brown, Stout, Dalessio, and Crosby (1988) estimated the correlation between biodata and task performance with a sample size of 16,230. (All other samples sizes were smaller than 10,000). When the Hough study is removed from the pairwise meta-analysis for conscientiousness and task performance, the results are very similar to those in Table 6: $r_c = .18$ with a corresponding standard deviation of .07. When the Brown et al. study is removed from its meta-analysis, the results are again very similar to the original findings: $r_c = .16$ and $SDr_c = .12$. In both cases, the significance tests still allow one to conclude that both effects are nonzero.

Next, I compared range restricted to unrestricted (or, more likely, less restricted) samples using the code that I assigned, as described in the Methods section. Based on the justifications for not correcting for range variation explicitly, I expected to see some correlations attenuated more than others. After removing the 25 samples that were considered to be unrestricted, the expected pattern was observed (see Appendix J for the correlation matrix pertaining to restricted samples). Therefore, range restriction is likely to have attenuated some correlations but in specific ways. Further work is needed to identify specifically in what settings and for which variable relationships attenuation (or enhancement) occurs.

Finally, I investigated discrepancies in the results of the pairwise meta-analyses and the multivariate meta-analysis. The pairwise results are necessarily flawed due to theoretical reasons discussed in the Methods section, regarding dependencies between correlations from the same sample. While I attempted to derive more accurate results with the multivariate method, the resulting estimates of population correlations revealed some strange patterns with eight correlations nearly equal to 1. As the multivariate method does not appear to have been thoroughly tested or applied in the research literature⁹, I explored some possible reasons for discrepancies with the pairwise results.

It was possible that discrepancies in results were due to sampling differences between the total sample and the subset of studies that were eligible for input into the multivariate analysis. I computed pairwise results (Table 12) for just the multivariate-

⁹ The most widely available article from the group, Raudenbush et al. (1988), has been cited just 10 times in the Social Sciences Citation Index as of July 2004.

eligible subsample, for comparison with the results in Table 6. It appears that some of the larger correlations became inflated to near 1 by the multivariate method but even these correlations are higher than expected (e.g., given the literature on scale intercorrelations between the Big Five personality dimensions). Given that two very different weighting schemes were used, the overall results do not vary too much apart from those correlations approaching 1.00 in the multivariate analysis. In any case, this difference does not explain the major discrepancies.

Another cause of the discrepancies could be that I incorporated statistical corrections into the multivariate method, corrections that had not been addressed by the original authors. I used corrected correlations as input and adjusted the computation of variances and covariances by the each sample's correction factor (i.e., measurement unreliability) because the standard error of a corrected correlation is larger than its uncorrected counterpart (Hunter & Schmidt, 2004). To evaluate this effect, I reran the multivariate analysis on the corrected correlations without correcting the covariances for attenuation. The results produced essentially the same patterns as the original multivariate analysis, discounting this as a reasonable explanation for discrepancies.

A third possibility is that certain samples received more weight and biased certain estimates upward. For example, three samples (i.e., one sample from Botwin & Buss, 1989, and two from Collins & Gleaves, 1998) contributed large correlations between extraversion and other personality variables (i.e., conscientiousness, agreeableness, and emotional stability), where most were larger than .7 before artifact corrections. I ran a multivariate analysis without these three samples and the corresponding ρ estimates were reduced to a "more reasonable" size (i.e., less than 1). However, further methodological

work needs to be done to determine whether 1) the results of those three studies, in this one exploration, were true outliers, 2) the multivariate method of weighting samples is inaccurate, 3) the method cannot handle values based on univariate corrections for measurement unreliability, or 4) sampling error produces large discrepancies between the pairwise and multivariate method when there are few studies.

Table 12

•1
<u>e</u>
Ē
Sa
e
ria
N
Iti
Ź
he
r th
s fo
tes
stimate
sti
(T)
Ilytic I
aly
an
sta-
Met
Se
Wi
air
ألم

		-	7	e	4	S	9	7	∞
(1)	Cognitive Ability								
6	Extraversion	0.							
3)	Conscientiousness	.07	.33						
(4)	Agreeableness	.01	.38	.63					
(5)	Openness	.23	.40	.16	.27				
9	Emotional Stability	.19	.46	.70	.63	.22			
6	Biodata	.32	.39	.51	.49	.53	.59		
(8)	Task Performance	.32	.03	.08	.04	.03	.08	.31	
6)	Citizenship Perf.	.17	.08	.19	.19	.08	.04	.54	.56

DISCUSSION

Review of Research Goals

The main goal of this research was to map the set of relationships between commonly studied and applied individual difference measures and mid-range concepts of job performance that are more detailed than just overall performance. This was accomplished by the creation of a meta-analytic matrix estimating the true correlations (or ranges of correlations for multiple populations with credibility intervals). The metaanalysis was intended to be comprehensive (within a specified time period) and generalizable, including all types of job samples that could be found in the literature. One drawback of such an approach is that estimated relationships showed considerable variance for the most part and additional moderator hypotheses were evaluated. This contradicts, in part, the proposition that processes related to citizenship behaviors are similar across jobs (Borman & Motowidlo, 1997).

Another major goal was to provide valid comparisons of relationships across meta-analyses to test the central proposition put forth by Motowidlo et al. (1997). Overall, the notion that task performance is predicted best by cognitive ability and citizenship is predicted best by other variables received partial support. Cognitive ability was a generally good predictor of both dimensions. Biodata are also good predictors but for reasons that are not entirely clear, perhaps because they capture cognitive, attitudinal, or other characteristics but probably not because they capture personality.

A specific contribution of this meta-analysis that adds to past work is the evidence that task and citizenship performance are related ($r_c = .49$; SD_{rc} = .36). Although this estimate shows a lot of variability, the credibility interval does not include 0, suggesting

that the range of population effect sizes, unmoderated, tends to be non-negligible. This has implications for theory but also for practical endeavors like the recent surge of work trying to identify how various predictor-criterion combinations can affect adverse impact on racial minority groups in personnel selection (e.g., Hattrup et al., 1998; Murphy & Shiarella, 1997; Schmitt et al., 1997).

At the same time, the estimated population correlation between task and citizenship performance is not necessarily as high as indicated in Figure 4. Halo error might be causing these strong observed relationships (Conway, 1996). Or, the relationship could be biased upwards because of selective sampling in primary studies. The people most likely to be studied in an experiment are job incumbents who are range restricted on task performance to some degree (i.e., they perform well enough to maintain their jobs), but who may be range enhanced if study participation is related to performing more citizenship behaviors and helping out an organization; the "normal" population would be range restricted on citizenship performance compared to experimental samples. *Summary of Findings*

Overall conclusions pertaining to the specific hypotheses are included in Table 13 with the specific results summarized below. Unfortunately, there were not many studies available to conduct thorough statistical tests of hypothesized moderators and differential prediction patterns. Small sample sizes are associated with low statistical power and susceptibility to second-order sampling error. Hypotheses 2, 3 and 4 could not be tested directly using traditional statistical methods as a consequence of the small sample sizes, while two others could not be tested at all (H11 and H12C). Additionally, the pattern of results is not always stable. Therefore, it may be premature to conduct a meta-analysis

Table 13

Summary of Conclusions for Hypotheses

Hypothesis	Conclusion
H1: Noncognitive predictors will have higher validities with citizenship for managerial / sales jobs.	Not Supported
H2: Citizenship dimensions of job dedication and interpersonal facilitation will produce differential patterns of validity.	Not Supported
H3: Task performance will be related more strongly to cognitive ability than to personality.	Supported
H4: Citizenship performance will be related more strongly to personality than to cognitive ability.	Not Supported
H5: Job complexity will moderate the relationship between cognitive ability and task performance.	Not Supported
H6: Job complexity will not moderate the relationship between cognitive ability and citizenship performance.	Not Supported
H7: Conscientiousness will be related to task and citizenship performance.	Supported
H8: Emotional stability will be related more strongly to citizenship than to task performance.	Not Supported
H9: Agreeableness will be related to citizenship performance only.	Supported
H10: Opennness to experience will be related to task performance only.	Not Supported
H11: Interviews will be related to task and citizenship based on what they measure.	Not Tested
H12: Biodata will be related to task and citizenship based on what they measure.	Partially Supported

such as this, despite the fact that smaller meta-analyses have been published on almost every portion of the correlation matrix created here (Table 5). In spite of everything, the results are useful because they represent the current state of research related to theories of citizenship performance and offer insight about where future research efforts can be focused. Furthermore, these results are not evident simply by surveying the literature or by using simple vote-counting methods, based on p-values.

With respect to Hypothesis 1, managerial and sales jobs moderated the validity of noncognitive measures with citizenship performance but in the opposite direction of that hypothesized. Managerial/sales jobs produced a stable validity (r_c) of .16. Although the validity of other jobs was much higher at .31, it was unstable with only 18% of the variance explained by artifacts. I hesitate to speculate at this point about why managerial/sales jobs produce "lower" validities than other jobs because it may be the case that one specific grouping of jobs is causing the "other" group to have validities higher than .16. However, the results support the idea that focused work on either managers (e.g., Conway, 1999) or salespeople (e.g., MacKenzie et al., 1991) might not generalize to other settings.

Regarding Hypothesis 2, a direct test could not be conducted due to small sample sizes. Other tests of individual effects suggest that job dedication and interpersonal facilitation produce essentially the same pattern of relationships across cognitive ability, personality, and task performance, and that they are highly correlated with one another. Still, many studies were confounded with common method variance and similar biases because they measured these two citizenship facets with the same instrument and rater. Overall, these findings tentatively support the job dedication and interpersonal facilitation

as facets of a single citizenship performance construct, allowing for the use of more parsimonious theories. This also leads to the practical conclusion that there is little need to validate predictors of citizenship separately for the two dimensions, as in Organ and Ryan (1995)'s meta-analysis.

As for Hypothesis 3, cognitive ability was the dominant predictor of task performance. This is unsurprising given the evidence for validity generalization and theories about performance. What is surprising is the conclusion for Hypothesis 4, that cognitive ability is also one of the best predictors of citizenship, not being outperformed significantly by any of the personality variables as was hypothesized by Motowidlo et al. (1997). The implications here are that cognitive ability is always useful and the advantages and disadvantages (e.g., adverse impact) associated with it cannot be avoided. Such conclusions have been made by others more generally (e.g., Sackett, Schmitt, Ellingson, & Kabin, 2001) or based on less compelling empirical evidence (e.g., Hattrup et al., 1998). It is still true that personality predictors can explain variance in job performance, but they are not likely to be good substitutes for cognitive ability, though conscientiousness comes close in predicting citizenship performance.

The results associated with Hypotheses 5 and 6 were puzzling as they seem to contradict previous literature (e.g., Hunter & Hunter, 1984). Job complexity was not found to moderate cognitive validities specifically with task performance. Still, the moderation found in previous research has pertained mostly to broad measures like overall performance, leaving open the possibility that nontask components of performance produce extra variance that is moderated by job complexity. Again, I cannot offer more than speculation for this null finding.

What is interesting, on the other hand, is the strong finding that job complexity moderates cognitive validities with citizenship. This formally contradicts Hypothesis 6 but that hypothesis was posed in contrast to H5, without much theory to guide it. One possible explanation for this finding is that intelligent people can finish core tasks and immediate responsibilities faster than others when the tasks are low in complexity. This might, in turn, lead to spare time and resources that are used to perform citizenship. For example, a coworker who has already completed his main tasks is more likely to help another than someone who has a backlog of work to finish. Another possible explanation is that complex jobs tend to provide greater opportunities for citizenship and all employees are expected to perform them, rather than just the ones with high cognitive ability. At some level, everyone may be expected to endorse the organization or to show personal initiative. However, the same effect might be hypothesized for low level jobs if situational influences were strong enough (e.g., in a Total Quality organization).

Hypotheses 7 through 10 involving direct estimates of correlations were easier to test than the previous hypotheses. Conscientiousness was the only strong personality predictor of task performance but it, as well as agreeableness, predicted citizenship relatively well.

Neither emotional stability nor openness was a good predictor of either performance dimension, as hypothesized. At the same time, extraversion significantly predicted citizenship performance, based on the multivariate results but not the pairwise results.

Together, the findings associated with the conscientiousness and agreeableness match the results of previous studies and meta-analyses on OCBs (e.g., Borman et al., 2001). The findings for the other three personality dimensions, however, differed from

past findings. One unique aspect of this meta-analysis that might explain some of the discrepancies is that a broader range of measures and settings were included here, whereas past meta-analyses studied specific groups (e.g., applied samples, managers, or salespeople). Generally, it seems plausible that a particular context can determine, at least in part, whether certain personality characteristics are helpful for performing citizenship (or task) behaviors. Employees who work individually are more likely to draw upon conscientiousness to improve their overall performance whereas employees in a social or team-based atmosphere can improve their contribution to the organization either through being conscientious or being more interpersonally helpful. Because observed personality validities have varied within and across meta-analyses, more controlled laboratory work may need to be done to isolate specific causal effects.

Hypothesis 11 could not be tested given the data collected but Hypothesis 12A suggests that the more biodata assess cognitive ability, the more valid it will become in predicting task performance. Given the results and explanations for Hypothesis 3, this finding is self-evident. Hypothesis 12B was also supported statistically but the size of the effect was nominal. The proportion of biodata assessing personality led to small increases in the correlation between biodata and citizenship. This suggests that characteristics other than personality will act through biodata to predict citizenship.¹⁰ Organ and Ryan (1995) made the strong conclusion that attitudinal variables were more effective in predicting citizenship. A thorough examination of attitudinal variables was beyond the scope of this study and must be relegated to future studies. Additionally, Hunter and Hunter (1984) also found biodata to have good validities across multiple

¹⁰ A review committee member noted that, because there was little variance regarding the proportion of personality assessed by the biodata measures, range restriction may have produced the null result.

types of performance criteria in their review of various performance predictors but argued that the operational validity of biodata might be considerably lower.

Future Directions

There was the general limitation of small sample sizes for many of the analyses here. As a result, I must recommend that additional work be carried out on all aspects of citizenship predictors since the findings here are not completely consistent with some other reviews like the meta-analysis of personality and OCBs by Borman et al. (2001). It is important to understand moderators of the different relational patterns across variables under more controlled conditions as meta-analytic moderator analyses cannot escape certain confounds when study characteristics or statistical artifacts covary with true moderators (Russell & Gilliland, 1995). And the results certainly suggest that there are moderators left to be identified.

Specific recommendations related to the research hypotheses evaluated here are as follows. The reasons as to why managerial and sales jobs produced more stable cognitive - citizenship validities than other jobs are unclear. And it seems that, either all other jobs produce higher validities on average or, more likely, that another group of as yet unidentified jobs produces very strong validities. Clearly, detecting what kinds of moderators do produce stable validities in other jobs would help increase our understanding of citizenship performance processes. Similarly, researchers should attempt to replicate and explain the finding that cognitive ability is related more strongly to citizenship in less complex jobs.

I concluded here that the findings for the job dedication aspect of citizenship were not distinct from findings for the interpersonal facilitation aspect. There is a strong

reason to believe that biases like halo or common method variance inflated the relationship between these dimensions to some degree. Future research should try to verify the extent to which this assumption holds. Although individual differences may not predict these different types of performance behaviors well, there is still a substantive distinction here from a content validity standpoint since organizations may be interested in increasing one type of behavior or the other.

The biodata results suggest that similar types of analyses (hopefully on larger data sets) can increase our knowledge about the predictive power of nonconstruct measures and that such measures may produce higher validities than component construct measures. Biodata are ambiguous and are applied inconsistently across settings, possibly measuring many different things. They are rarely said to be drawing upon distinct constructs, though some scales of the ABLE have been considered measures of personality constructs (Bobko et al., 1999). This seems to have caused some to shy away from theory-based biodata. This study and past work (e.g., Hunter & Hunter, 1984; Schmitt et al., 1997), however, suggests that there are significant practical benefits associated with biodata use in predicting performance. Clearly, there is a need to go beyond this examination of personality and cognitive constructs to determine what other aspects of biodata help to predict outcomes. New findings may actually help others to develop biodata that are more construct oriented. Also, measures of attitudes or situational influences like those studied by Organ and Ryan (1995) may be related to molar measures like biodata and may show more promise for predicting and understanding citizenship behaviors.

Future work can also address other issues that were not investigated here but that have been applied in the past or to other types of performance measures. Researchers have consistently found differences between various types of measurement methods. Meta-analyses conducted by Ford et al. (1986) and Bommer, Johnson, Rich, Podsakoff, and MacKenzie (1995), show that there are meaningful statistical differences between subjective and objective measures. Podsakoff et al. (2000) suggested that examinations of multiple performance criteria like citizenship and task behaviors might be influenced by rater biases and common method variance. Such influences would artificially inflate correlations.

If the correlation between task and citizenship performance is as high as it is estimated to be here, however, there are implications for the trend of research on reducing adverse impact for racial minority groups by using various weighting schemes of predictors and criteria in selection (e.g., De Corte, 1999; Hattrup et al., 1998; Murphy & Shiarella, 1997). These schemes are only meaningful inasmuch as the multiple variables entered into a model are providing unique information. If measures of task and citizenship performance provide redundant information in practice, these complex methods of predicting performance will be less effective.

Another potential area of research would be the examination of the accuracy in measurement of citizenship behaviors since they are, almost by definition, more abstract and difficult to notice, especially if considered to be "extra-role" (Turnipseed, 2002). For example, Chen and Francesco (2003) included the item: "Complies with company rules and procedures even when nobody watches and no evidence can be traced." This item begs the question of whether supervisors can or will notice certain acts of citizenship.

Lovell et al. (1999) found that men and women received similar ratings for overall performance despite women having a higher likelihood of performing citizenship behaviors. Thus, researchers might obtain more accurate results by accounting for differences in measurement of task and citizenship performance, apart from statistical considerations of (e.g., of reliability).

Finally, I was concerned with the middle-range distinction between task and citizenship performance and focused on the salient issues regarding how citizenship is conceptualized and studied to this end. This endeavor was also motivated in part by the large body of more recent work on noncognitive predictors and the assumption that they are primarily beneficial for understanding citizenship. However, there are many different forms of task performance that might moderate the results found here. It is true that cognitive ability already tells us much about task performance but this meta-analysis suggests that aspects of task performance may be strongly related to citizenship performance. (It also seems that conscientiousness does not predict task performance well, as has been implied by past validation work.)

Although unrelated to theory testing, future methodological and pedagogical efforts are needed to make multivariate meta-analytic methodologies useful. Results of the supplemental analyses offer some explanation for discrepancies between the traditional pairwise approach and the multivariate approach but neither could be said to be perfectly accurate. From the perspective of testing "sensitivity" the two analyses show what kinds of results can be derived from the data given different sets of assumptions. Consequently, there is still a need to identify the advantages and disadvantages of using

the multivariate method with real world data if users are to understand whether results are more or less accurate than those produced by more traditional methods.

Limitations

This meta-analysis was limited in several ways which caused some of the conclusions to be ambiguous until additional research is available. First, the number of studies available for estimating each bivariate relationship varied greatly and was quite small in some cases. Where the sample size (k) did not preclude a meta-analysis (i.e., as it did with the structured interview), the power to detect moderators was relatively low (Hedges & Pigott, 2001). Artificially small cells are unrepresentative of the population of studies and can contain wildly inaccurate estimates due to second order sampling error (i.e., error due to a small number of primary studies); effect sizes can be very different or very similar solely due to chance. The mean estimates of ρ however are relatively unaffected by the number of studies when the average sample size (N) within each study is large, as was the case here.

Second, the quality of estimates from primary studies (Lipsey & Wilson, 2001) were believed to be of good quality (having been conducted by many well-respected researchers), but the accuracy of meta-analytic estimates depends on the accessibility of accurate information. These results might be biased upward to some degree for reasons related to sampling studies including but, not limited to, the effects of publication bias towards significant results or well written research, the file drawer effect where studies with null findings are not published, and the exclusion of dissertations which tend to be of weaker quality (Ashworth, Osburn, Callender, & Boyle, 1992; Campbell, Dunnette, Lawler, & Weick, 1970; Rotton, Foos, Van Meek, & Levitt, 1995). Hunter and Schmidt (2004) show that "missing a few studies randomly usually does not reduce the accuracy of a meta-analysis by nearly as much as might be supposed" (p. 85). Nonetheless, this meta-analysis is limited by how well the sampled studies represent the actual universe of true relationships.

Another limitation of this study is related issues of measurement that come into play when conducting a meta-analysis or primary study. My aim was to examine the conceptual relationships between multiple individual difference variables to predict job performance. The type of method used to measure those variables can create bias or be susceptible to bias, particularly if they are subjective measures (e.g., Allen & Rush, 2001). The low agreement for coders' ratings of measure subjectivity prohibited a specific test of this moderator but there was a tendency for citizenship to be measured more subjectively. If future research establishes such a link, there would be a number of obvious implications for the use of different performance dimensions in research and practice, based on research already mentioned.

The study is also limited in that a statistical artifact known to be operating could not be corrected: range variation. Because many of the samples were determined to be range restricted, the results are probably attenuated. However, it is not clear whether all results are attenuated uniformly or some are attenuated while others are enhanced. I imagine that the latter condition is true. It is worth mentioning that range enhancement, which has been largely ignored when studying cognitive ability, may occur in the study of citizenship behaviors. Becker and Randall (1994) found that employees who returned an attitude survey also performed more citizenship behaviors than their nonrespondent counterparts. The implication of such a phenomenon for this study is that correlations

will be biased upwards. There is a need for additional work on this topic before one can make accurate statistical corrections to the data if some relationships are range restricted while others are range enhanced.¹¹

Also, this study is relevant only insofar as it accurately categorized the findings in primary studies in a meaningful way. The definitions of variables used to classify studies, particularly for the newer performance concepts (i.e., citizenship, job dedication, and interpersonal), were based on a broad body of literature and should exhibit acceptable face validity. I also attempted to categorize studies according to these definitions accurately using coding rules. In this study, the number of categorization errors was noticeably high (i.e., the accuracy rate was low). The index was imperfect and interrater agreement is not equivalent to capturing true scores but the agreement results suggest that more refined definition and measures of citizenship would be helpful in future work. Future syntheses should attempt to refine the definitions used here, and to apply them more accurately in classifying the results of primary studies.

Overall Conclusion

The results of this study partially support the theory proposed by Motowidlo, Borman, and Schmit (1997). Task and citizenship appear to be moderately correlated but distinct aspects of job performance. Although different performance predictors seem to show differential validity between the two dimensions of performance as theorized, cognitive ability is the single most effective predictor across the two dimensions. It does appear to be the case that personality dimensions can predict aspects of citizenship behavior but more research is needed to understand which aspects of personality, when,

¹¹ What determines whether a sample is range restriction and enhancement is its range compared to the range of a reference group to which one wishes to generalize findings.

and how they are important. Alternatively, biodata were good predictors of both task and citizenship performance. Biodata validities were not greatly affected by the degree to which various constructs were assessed. Although not all hypotheses could be fully tested quantitatively, the patterns of results have implications for past research on validity and adverse impact, as well as future theoretical work on citizenship performance determinants.

Appendix A

Pilot Coding Sheet

For any N/A value, enter 9090

Study Descriptives

1. Study ID

(If a study reports multiple independent studies with distinct outcomes and samples, add a **decimal** to the Study ID and code each study **separately**.)

- 2. First Author (last name first):
- 3. Year (last two digits):

(If multiple reports of the same study, then code year of the more "formal" publication.)

- 4. Published: Yes/No
- 5. Reference source:

 book
 journal article
 book chapter
 thesis or dissertation
 technical report
 conference paper
 other : ______
- 6. Citation (APA form):

Study Sample

- 7. Type of population sampled (as described in paper):
- 8. Study type:
 - 1 Predictive
 - 2 Concurrent
- 9. 1 Applied setting 2 Experimental
- 10. 1 US
 - 2 European
 - 3 Other:
- 11. Total sample size (usable cases)

12. Mean age of sample (at start of study)

13. Job incumbents or job applicants

<u>Racial/Gender Breakdown</u> – Fill in whatever is reported by the study (either % or N)

- 14. White *N*
- 15. White %
- 16. Black *N*
- 17. Black % ______ 18. Hispanic N
- 19. Hispanic %
- 20. Asian N
- 21. Asian %
- 22. Males *N*
- 23. Males %

24. Official jobs included and # of people in each group (list)

This section to be coded separately using information obtained in #24. (circle all that apply)

Ν

- 25. Managerial Nonmanagerial Other
- 26. Attrition: (N) 1 Refuse to complete study
 - 2 Quit study
 - 3 Thrown out by researcher
 - 4 Other: ______ Reason?:
- 27. Reason for missing data?

28. Study design

1 Predictive

2 Concurrent

29. Was there a manipulation in between IV & DV measures? If so, what?

30. Are performance measures considered DV's in study? Yes/No Item 31 to be coded separately using information obtained from ONET.

31. Job complexity: The amount of information, knowledge, and concepts that must be dealt with for regular job tasks (Avolio & Waldman, 1990).

1 Low (e.g., line workers)

2 Medium (e.g. supervisors, skilled workers)

3 High (e.g. professionals, specialists, upper management)

Task Performance

- 32. Label used by authors:
- 33. Operationalization/Definition:
- 34. Constructs thought to be measured:
- 35. Breadth of measure: (does it measure a narrow aspect or specific task?)
 - 1 Broad
 - 2 Narrow
- 36. Measure used: _____

Established reliability:

<Attach a photocopy of items if they are provided>

37. Who makes the ratings?

- 1 Self-report
- 2 Peer
- 3 Supervisor
- 4 Other _____
- 38. Range Restricted?

Due to selection _____ (check if yes) List other possible reasons and any estimates of the magnitude

- 39. Training or job criterion
 - 1 Training
 - 2 On-the-job

Citizenship Performance

- 40. Label used by authors:
- 41. Operationalization/Definition: (for qualitative purposes)
- 42. Constructs thought to be measured: (for qualitative purposes)
- 43. Nature of acts: (for moderator analysis)
 1 Self-discipline (see definition)
 2 Interpersonal (")
- 44. Breadth of measure: (does it measure a narrow aspect or specific task?) 1 Broad
 - 2 Narrow
- 45. Measure used: _____

Established reliability:

<Attach a photocopy of items if they are provided>

- 46. What is measured?1 behavioral frequency2 behavioral quality
- 47. Who makes the ratings?
 - 5 Self-report
 - 6 Peer
 - 7 Supervisor
 - 8 Other
- 48. Range Restricted? Due to selection _____ (check if yes) List other possible reasons and any estimates of the magnitude
- 49. Training or job criterion
 - 1 Training
 - 2 On-the-job

Structured Interview

- 50. Label used by authors:
- 51. Operationalization/Definition:
- 52. Constructs thought to be measured:
- 53. Breadth of measure: (does it measure a narrow aspect or specific task?) 1 Broad 2 Narrow
- 54. Measure used:

Established reliability:

<Attach a photocopy of items if they are provided>

- 55. Structured or unstructured: majority of questions specified beforehand
- 56. How many raters (one vs. panel)?
 - 1 Single person
 - 2 Multiple raters
 - 3 Panel:
- 57. Who makes the ratings?
 - 9 Self-report
 - 10 Peer
 - 11 Supervisor
 - 12 Other
- 58. Range Restricted? Due to selection _____ (check if yes) List other possible reasons and any estimates of the magnitude
- 59. Amount of structure: (Based on Huffcutt & Winfred, 1994; Huffcutt & Roth, 1998)
 - 1 Low - standardization of topical areas to be covered
 - 2 Medium
 3 High
 at least half of the questions are specified beforehand
 majority of questions specified

Biodata

- 60. Label used by authors:
- 61. Operationalization/Definition:
- 62. Constructs thought to be measured:
- 63. Breadth of measure: (does it measure a narrow aspect or specific task?)
 - 1 Broad
 - 2 Narrow
- 64. Measure used: _____

Established reliability:

<Attach a photocopy of items if they are provided>

- 65. Scoring Key:
 - 1 Empirical
 - 2 Rational
 - 3 Factor-analytic keying
- 66. Type of items included:
 - 1 Strictly biodata (objective measures of past experiences)
 - 2 Attitudinal
 - 3 Hypothetical or future events
- 67. Who makes the ratings?
 - 13 Self-report
 - 14 Peer
 - 15 Supervisor
 - 16 Other _____
- 68. Range Restricted?

Due to selection _____ (check if yes)

List other possible reasons and any estimates of the magnitude

Cognitive ability

- 69. Label used by authors:
- 70. Operationalization/Definition:
- 71. Constructs thought to be measured:
- 72. Breadth of measure: (does it measure a narrow aspect or specific task?)1 Broad2 Narrow

.

73. Measure used: _____

Established reliability:

<Attach a photocopy of items if they are provided>

- 74. Who makes the ratings?
 - 17 Self-report
 - 18 Peer
 - 19 Supervisor
 - 20 Other _____
- 75. Range Restricted?

Due to selection _____ (check if yes) List other possible reasons and any estimates of the magnitude

Personality

- 76. Label used by authors:
- 77. Operationalization/Definition: Of personality:

Of each dimension:

78. Construct measured (circle all that apply):

- 1 Conscientiousness
- 2 Extroversion
- 3 Agreeableness
- 4 Emotional Stability
- 5 Openness to experience / Intellectance
- 6 Positive Affectivity
- 7 Negative Affectivity
- 8 Other (specify label): _____
- 79. Breadth of measure: (does it measure a narrow aspect or specific task?)
 - 1 Broad
 - 2 Narrow
- 80. Measure used: _____

Established reliability:

<Attach a photocopy of items if they are provided>

- 81. Who makes the ratings?
 - 21 Self-report
 - 22 Peer
 - 23 Supervisor
 - 24 Other _____

82. Range Restricted?

Due to selection _____ (check if yes)

List other possible reasons and any estimates of the magnitude

Miscellaneous

Fill in the matrix with the data that are provided. If more than one estimate is given, explain this below and put the average (sample-size weighted) estimate in the shaded region. Place reliability estimates on the diagonal.

	1	2	3	4	5	9	7	8	6	
1. Task										
2. Citizen										
3. Cognitive										
4. Conscientiousness										
5. Biodata										
6. Interview										
7. Agrecableness										
8. Extroversion										
9. Emotional Stability										
10. Openness										

84. Reliability coefficient used:

Alpha
 Interrater / Parallel-forms
 Test-retest
 Assumed
 Other: ______

If different ones are used for different measures, explain which coefficients match with the variables:

85. Why there is more than one estimate for a given relationship:

Calculations for obtaining average estimate:

	Overall Mean	SD	Type of Reliability	Objective	White- Black	White- Hispanic	White- Asian	Male- Female
1. Task								
2. Citizen								
3. Cognitive								
4. Conscientiousness								
5. Biodata								
6. Interview								
7. Agreeableness								
8. Extroversion								
9. Emotional Stability								
10. Openness								

Definition List

Task performance: Behaviors that directly or indirectly (through other workers) affect core production that transforms input to output or delivers a service.

Citizenship performance: Behaviors that 1) are not directly related to core tasks and 2) support the social and/or psychological environment. Examples include: loyalty, cooperative behaviors (not affecting core production), whistle-blowing, sportsmanship, prosocial behavior, personal initiative, showing extra effort and perseverance, volunteering to do extra or unrelated work. Counterproductive or retaliatory behaviors are NOT included.

Self-discipline: Citizenship behaviors that do not require a direct interaction with another person. Examples: working hard, taking initiative, and following organizational rules.

Interpersonal: Citizenship behaviors that require a direct (not necessarily face-to-face)

interaction with another person.

Structured interview: A structured interview, at the very least, evaluates a response to each question posed to the interviewee (from Huffcutt & Roth, 1998)

Biodata: A measure of background life experiences that is intended to predict future behaviors of the same type.

Cognitive ability: Broadly, any test of computational, problem solving, or mental abilities.

Personality: Enduring characteristics of the individual that are tied to one of the Big Five dimensions: Conscientiousness, Agreeableness, Extroversion, Emotional Stability, and Openness to Experience. A measure may capture smaller aspects of any one dimension but not overlap with another dimension.

[From Barrick, Mount, & Judge (2001)]

Conscientiousness: dependability, achievement striving, and planfulness.

Extraversion: sociability, dominance, ambition, positive emotionality, and excitement-seeking.

Agreeableness: cooperation, trustfulness, compliance, and affability.

Openness to experience: intellectance, creativity, unconventionality, and broad-mindedness.

Emotional stability: lack of anxiety, hostility, depression and persona insecurity.

	-
8	*
	ð.
×	<u> </u>
•	Ę.
P	S
2	50
ē	
ā.	<u> </u>
5	÷
<u> </u>	×
	2
-	\mathbf{O}

Coder:
y code:
Study

Citation:

Source Type (1=Journal, 2=Book chapter, 3=Book, 4=Other):

The sample: Kind of job/workers (Description) 1=Manager 2=Nonmanagerial 3=Both	O*NET code:		
<u>Total number of samples studied</u> <u>paper:</u> Sample sizes	id in this		
Design Predictive? (1=yes, 2=concurrent) [] IF predictive, how much time between measures?:) [
Missing data? Summary of reason given. Manipulation between predictor	r and performance?		

¹ This is actually a formatted version of the coding sheet. Coding sheets were made and used in Microsoft Excel. The content collected is the same but some of the extra wording and examples are removed for ease of viewing. Cells are smaller than normal size. No essential information has been omitted.

	5												
	Cog	Extra		Consc Agree	Open	E. Stab	Bio	Interv	Task	Ctzn	J. Dedic Interp	Interp	Ovr
Measure													
Broad? (B or N)													
Subjective?													
Reliab number: Reliability type:			_										
Sample 1*						CORRELATIONS	LATIO	NS		1			
	g Ex	tra.	Cog Extra. Consc.	Agree	Open	Open E. Stab.	Bio	Interv	Task	Ctzn		J. Dedic Interper	erp
Cognitive				1									
Extraversion													
Conscient.													
Agree													
Open		-											
E. Stability													
Biodata													
Interview													
Task													
Citizenship		-											
Dedication											Chillen of		
Interpersonal													
Overall													

Appendix B (continued)

139

Appendix C Code Book

Preparation:

- Find the template coding file. Make a copy of it for each article.

Appendix C

(continued)

Definition List

<u>Task performance</u>: Behaviors that directly or indirectly (through other workers) affect core production that transforms input to output or delivers a service. Sometimes these are referred to as "in-role" behaviors because they are tied to one's job roles (but if the role includes non-task behaviors, then the two concepts may be very different).

<u>Citizenship performance</u>: Behaviors that 1) are not directly related to core tasks and 2) support the social and/or psychological environment. Examples include: loyalty, cooperative behaviors (not affecting core production), whistle-blowing, sportsmanship, prosocial behavior, personal initiative, showing extra effort and perseverance, volunteering to do extra or unrelated work. Counterproductive or retaliatory behaviors are NOT included.

- <u>Self-discipline</u>: Citizenship behaviors that do not require a direct interaction with another person. Instead, they are related to helping the organization overall. Examples: working hard, taking initiative, and following organizational rules.
- <u>Interpersonal</u>: Citizenship behaviors that require a direct (not necessarily faceto-face) interaction with another person. Examples: helping others and backing people up.

<u>Structured interview</u>: A structured interview, at the very least, evaluates a response to each question posed to the interviewee (from Huffcutt & Roth, 1998)

<u>Biodata</u>: A measure of background life experiences that is intended to predict future behaviors of the same type.

<u>Cognitive ability</u>: Broadly, any test of computational, problem solving, or mental abilities.

<u>Personality</u>: Enduring characteristics of the individual that are tied to one of the Big Five dimensions: Conscientiousness, Agreeableness, Extroversion, Emotional Stability, and Openness to Experience. A measure may capture smaller aspects of any one dimension but not overlap with another dimension.

<u>Conscientiousness</u>: dependability, achievement striving, and planfulness. <u>Extraversion</u>: sociability, dominance, ambition, positive emotionality, and excitement-seeking.

<u>Agreeableness</u>: cooperation, trustfulness, compliance, and affability. <u>Openness to experience</u>: intellectance, creativity, unconventionality, and broad-mindedness.

<u>Emotional stability</u>: lack of anxiety, hostility, depression and persona insecurity.

Appendix C

(continued)

Using the excel file:

- 1) Enter your initials on line 3 next to "Coder"
- 2) For "citation" write the article citation in APA format.
- 3) For "Source Type" enter the appropriate response
- 4) Go to the methods section of the article and briefly summarize the type of people studied.
- 5) Use your judgment based on the description to decide whether the jobs are managerial or not, or including both types. Managerial jobs usually involve the supervision of other workers.
- 6) How many studies or samples are included in the article? The main thing to look at is how many different groups of people were analyzed. Answer on line 13.
- 7) Give the size of each sample/group studied. Answer on line 14, use other additional columns if necessary.
- 8) Is this study about people applying for a job, who are later evaluated on job performance? If so, then choose 1 for predictive. If the study measures current workers, then select concurrent.
- 9) For line 18, answer only if line 17 was predictive. How much time passed between the time when the predictors were measured (interview, biodata, personality, cognitive ability) and job performance. If the length of time is different depending on the predictors, then note that (e.g., for cognitive ability, 3 weeks).
- 10) When the sample size is given in the article, sometimes not everyone was used in the analyses. If there were a lot of people who didn't provide useable data, there might be some description of this in the text. You should also check the tables/correlation matrix when you get correlations to see what the sample size is (N=...), to make sure this matches with the number you put in line 14.
- 11) Sometimes a study will measure the predictors and then manipulate the workers (e.g., give them a training session). If this is the case, note what happened in line 23. If there manipulation is not <u>between</u> one of the predictors we care about and job performance, then don't mention this.
- 12) There are two types of measures: predictors (cognitive, personality factors, biodata, and interview) and job performance (overall, task, or citizenship). Citizenship can be broken into two smaller categories (see definition list above).

Find the measure used to assess the variables and write it in the corresponding box on line 28.

Appendix C

(continued)

- 13) Use your judgment to choose whether the measure is broad or narrow, based on the definition list. For example, if cognitive ability is measured with a short math test, that would be narrow because it captures a specific aspect of cognitive ability. Or, if task performance is measured by the number of people called per week, that might be narrow if you think there is more to performance for a sales job.
- 14) Again, use your judgment here to decide whether the measure is objective or subjective. Objective measurements should be similar no matter who is providing the data (whether the manager is rating a worker's performance or a worker is taking a cognitive test). If the measure asks about self-reported feelings or perceptions and depends on the situation or time, then it would be subjective (e.g., most personality measures).
- 15) Find the reliability for the measure used. If it's a single rating like overall job performance by a supervisor, then it probably won't be listed. For line 31, write the number. For line 32, write down what type of reliability estimate the number means: alpha coefficient/internal consistency, interrater (between multiple raters who are measuring the same thing), test-retest (at different times). If you cannot tell or it isn't mentioned, write that down. Sometimes, these numbers will be on the "diagonal" line of the correlation matrix (check the footnote to make sure).
- 16) Find the correlations. Usually, these will be in correlation matrix and you won't need to look through the text. If they are presented in more than one table, find out why. If there are multiple tables because there are different samples, use the other excel spreadsheets (in the same file). If there are multiple tables because the article authors are describing the same people in different ways, put all the numbers in the same table but label why they are different.

Appendix D Interrater Coding Agreement Results

Data are provided for the four (italicized) categorical ratings by the author and undergraduate assistant made on the coding sheet. N is the number of ratings made for the 29 published studies. Ratings that could not be made based on the information given were also treated as a category called "Undetermined."

Undergrad.	Manager Rating Graduate StudentManagerM0.15N0.06B	N=33 <u>Nonmanager</u> 0.06 0.70	<u>Both</u> 0.03
Undergrad.	Predictive Rating Graduate Student PredictiveP0.03C0.09U0.09	N=33 <u>Concurrent</u> 0.03 0.67 0.06	<u>Undetermined</u> 0.03
	Broad Rating		
Undergrad.	Graduate Student Broad B 0.70 N 0.02 U	N=105 <u>Narrow</u> 0.15 0.02	Undetermined 0.10
Undergrad.	Subjective Rating Graduate Student Subjective S 0.58 O 0.20 U	N=125 <u>Objective</u> 0.06 0.06	Undetermined 0.10

SAS / IML program for multivariate computations.

The following code for this paper was generated using SAS software, Version 8.02 of the SAS System for Windows. Copyright © 2001 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

The syntax below for a level 2 regression only works with two moderators and three correlations.

*Input file of correlations, tab delimited; data thesis; infile 'c:\file1.txt' recfm=v dsd dlm='09'x lrecl=5000; input cogcon cogagr cogB CogTask CogDed CogInterp conagr conB ConTask ConDed ConInterp agrB AgrTask AgrDed AgrInterp BTask BDed BInterp TaskDed TaskInterp DedInterp; run;

*Input file of correction factor associated with each correlation; data thesisc; infile 'c:\file2.txt' recfm=v dsd dlm='09'x lrecl=5000; input cl c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 c13 c14 c15 c16 c17 c18 c19 c20 c21; run;

*Input file with sample size and Level 2 predictors; data thesism; infile 'c:\file3.txt' recfm=v dsd dlm='09'x lrecl=5000; input ss m1 m2; run;

proc iml; use thesis;	
read all into x;	*column vector of all correlations;
dat=x[,];	
use thesisc;	
read all into x2;	
correct=x2[,];	*matrix of correction factors;
use thesism (keep=m1);	
read all into x3;	
moderat=x3[,];	<pre>*matrix of moderators;</pre>
use thesism (keep=ss);	
read all into x3;	
sample=x3[,];	*sample size vector;
sigma=.;	*Sigma column vector;

(continued)

n=nrow(sample);	*number of cases;
numr=ncol(dat);	*number of unique correlations inputed;
numrtot=n*numr;	*all correlations in entire meta-analysis;
nmod=ncol(moderat);	*number of moderators/predictors;

chart={2 1,3 3,4 6,5 10,6 15,7 21,8 28,9 36,10 45,11 55};

```
**** Variance module ****;
resultv=0;
start vari(r1,sam) global(resultv);
resultv=(1-r1##2)##2/sam; *Becker 2000, Eq. 4;
finish;
**** Covariance module ****;
resultc=0;
start covar(r1,r2,r3,r4,r5,r6,sam) global(resultc);
resultc=(.5*r1*r2*(r3##2+r4##2+r5##2+r6##2)+r3*r6+r4*r5-
(r1*r3*r4+r1*r5*r6+r3*r5*r2+r4*r6*r2))/sam; *Becker 2000, Eq. 5;
```

```
finish;
```

```
*Create Xmatrix and joint matrix with Level 2 predictors for calculations, only good for
2 predictors;
xmatrix=j(1,numr,0);
xmod=j(1,nmod,0);
current=j(1,numr,0);
current2=j(1,nmod,0);
do i=1 to n;
       do j=1 to numr;
               if dat[i,j] \diamond. then do;
                      current[j]=1;
                      xmatrix=xmatrix // current;
                      if j < nmod + 1 then do;
                              current2[j]=moderat[i,j];
                              xmod=xmod // current2;
                      end;
                      else xmod= xmod // \{0 0\};
                      current=j(1,numr,0);
                      current2=j(1,nmod,0);
               end;
       end;
end;
el=nrow(xmatrix);
xmatrix=xmatrix[2:e1,];
```

```
(continued)
```

```
e2=nrow(xmod);
xmod=xmod[2:e2,];
xmod=xmatrix || xmod;
**Make the var-cov matrix for whole data set**;
do i=1 to n:
       ** Row vector of correlations;
       caser = .:
       do j=1 to numr;
               if dat[i,j] \diamond. then caser = caser || dat[i,j];
       end:
       e1=ncol(caser);
       crow=caser[,2:e1];
                                            *Final vector;
       ** Row vector of correction factors;
       caser = :;
       do j=1 to numr;
               if correct[i,j] \Leftrightarrow. then caser = caser || correct[i,j];
                                                                          end;
       el=ncol(caser);
       acrow=caser[,2:e1];
                                     *Final vector;
       ****Add correlations to sigma vector ****;
       sigma = sigma // crow`;
                                  ******
       *****
*With the row of correlations for study(i), make a var-cov matrix*;
       *Prepare data in matrices to calculate covariances;
       isize=ncol(crow);
       pmat=j(isize,isize,1);
                                     *Initialize position matrix;
       pmat2=j(isize,isize,1);
       vcv=j(isize,isize,0);
                                     *Var-Cov matrix to be filled in;
       *Determine # of study variables based on # of input correlations;
       orig=0;
       if isize=1 then orig=1;
       else do q=1 to 10;
                if isize=chart[q,2] then orig=chart[q,1];
               end;
if orig=0 then print "Error for study" i "Data:" crow acrow isize orig;
```

*Place correlations for study(i) into a matrix; add=1; do j=1 to orig-1;

(continued)

```
do k=j+1 to orig;
                     pmat[j,k]=crow[,add];
                     pmat[k,j]=crow[,add];
                     add=add+1;
             end;
      end;
      *Create another matrix for correction factors;
      add2=1;
      *Place correction factors for study(i) into a matrix;
      do j=1 to orig-1;
             do k=j+1 to orig;
                     pmat2[j,k]=acrow[,add2];
                     pmat2[k,j]=acrow[,add2];
                     add2=add2+1;
             end;
      end;
if isize > 1 then do;
      **Make a list of codes for ordering correlations in study(i)**;
      minicycle=j(isize,2,1);
      order=1;
      do bs=1 to orig-1;
             do bt=bs+1 to orig;
                     minicycle[order,1]=bs;
                     minicycle[order,2]=bt;
                     order=order+1;
             end;
      end;
      **Compute all covariances between correlations in study(i)**;
      mcyclist=1;
      ps=1;
      pt=1;
      pu=1;
      pv=1;
      do j=mcyclist to orig-1;
             do k=j+1 to orig;
                     ps=minicycle[j,1];
                     pt=minicycle[j,2];
                     pu=minicycle[k,1];
                     pv=minicycle[k,2];
                     *Correction factor for each correlation involved;
```

(continued)

run

covar(pmat[ps,pt],pmat[pu,pv],pmat[ps,pu],pmat[ps,pv],pmat[pt,pu],pmat[pt,pv],sample[i
]);

```
cf=pmat2[ps,pt]*pmat2[pu,pv];
                            check=resultc/cf;
                                   print check;
                                   if check < 0 then check=0.0000001; *Changes
negative variance to near 0... it won't estimate 0;
                                   vcv[j,k]=check;
                                   vcv[k,j]=check;
              end;
       end;
 end;
       **Create variances in vcv**;
       do j=1 to isize;
              run vari(crow[,j],sample[i]);
              check2=resultv/acrow[,j]##2;
              if check 2 < 0 then check 2=0.0000001;
              vcv[j,j] = check2;
       end;
       ****Concatenate matrix for study(i) with S-matrix****;
       if i=1 then smatrix=vcv;
       else do:
                                                 *a and b are square matrices;
              addold = nrow(smatrix);
              addnew = nrow(vcv);
              if addnew = 1 then do;
                     new=j(addold,1,0);
                     old=j(1,addold,0);
              end;
              if addnew > 1 then do;
                     new=j(addold,addnew,0);
                     old=j(addnew,addold,0);
              end;
              smatrix = smatrix || new;
              bottom = old \parallel vcv;
              smatrix= smatrix // bottom;
       end;
                                   ******
```

end;

*Dump first missing value from sigma;

```
(continued)
```

e1=nrow(sigma); sigma=sigma[2:e1,]; *****Calculations*****; rho=inv(xmatrix`*inv(smatrix)*xmatrix)*xmatrix`*inv(smatrix)*sigma; rhov=inv(xmatrix`*inv(smatrix)*xmatrix); q=sigma`*(inv(smatrix)inv(smatrix)*xmatrix*inv(xmatrix`*inv(smatrix)*xmatrix)*xmatrix`*inv(smatrix))*sigm a; print rho; print rhov; print q; *****Levels analysis*****; *Level 2 regression paramater estimates; beta=inv(xmod`*inv(smatrix)*xmod)*xmod`*inv(smatrix)*sigma; *Var-cov matrix for parameters; vbeta=inv(xmod`*inv(smatrix)*xmod); *Test of overall model fit; he=(sigma-xmod*beta)`*inv(smatrix)*(sigma-xmod*beta); *Test of Model significance; hr=sigma`*inv(smatrix)*sigma - he; *parameter estimates, intercepts & slopes; print beta; print he; print hr; quit;

Appendix F Full List of Studies in Database

Indented citations are included within the parent citation. Citations with more than three authors have been abbreviated. Full citations can be found in the reference section.

Ackerman & Kanfer 1993 Allen & Rush 1998 Allworth & Hesketh 1999 Antonioni & Park 2001 Barbuto et al. 2003 Barksdale & Werner 2001 Barrick & Mount 1993 Barrick, Mount, Strauss 1993 Barrick, Stewart & Piotrowski 2002 Beaty, Cleveland & Murphy 2001 Becker & Vance 1993 Bell & Kozlowski 2002 Bell & Mengue 2002 Borman, White & Dorsey 1995 Borman et al. 1991 Bosshardt et al. 1992 Botwin & Buss 1989 Boudreau et al. 2001 Brown et al. 1988 Burroughs & White 1996 Caldwell & Burger 1998 Caliguiri 2000 Campion et al. 1988 Campion, Campion & Hudson 1994 Chad et al. 1999 Charbonneau & Nicol 2002 Chen & Francesco 2003 Collins & Gleaves 1998 Conway 1996 Boruch et al. 1970 Dickinson & Tice 1973 Forsythe et al. 1986 Gunderson & Nelson 1966 Gunderson & Ryan 1971 Holzbach 1978 Kavanagh et al. 1971 King et al. 1980 Lance et al. 1992 Lawler 1967

Orpen 1973 Tucker et al. 1967 Conway 1999 Cortina et al. 1992 Cortina et al. 2000 Baehr & Froemel 1977 Berkley 1984 Gully et al. 1998 Gully et al. 2000 Delery et al. 1992 Dicken, 1969 Dipboye et al. 1990 Exxon 1974 Exxon 1978 Friedland 1976 Friedland 1980 Lopez 1966 Motowidlo & Schmit 1996 Phillips & Gully 1997 Reeb, 1969 Roth & Campion 1992 Schmit 1996 Tubiana & Ben-Sitakhar 1982 Tziner & Dolan 1982 Costa & McCrae 1988 Crant 1995 Cropanzano, Rupp & Bryne 2003 Dalessio & Silverhart 1994 Day & Silverman 1989 De Fruyt & Mervielde 1999 Deadrick & Madigan 1990 Douthitt, Eby & Simon 1999 Farh, Podsakoff & Organ 1990 Farh, Werbel, & Bedeian 1988 Ferris, Witt & Hochwarter 2001 Findley, Giles & Mossholder 2000 Gellatly & Irving 2001 Gellatly 1996 George 1991

Appendix F

(continued)

Goffin, Rothstein & Johnston 1996 Hansen 1989 Hattrup, O'Connell & Wingate 1998 Haworth & Levy 2001 Hedge & Teachout 1992 Hochwarter, Witt & Kacmar 2000 Hogan, Hogan, Gergory 1992 Hogan et al. 1998 Hough 1992 Hough et al. 1990 Huffcutt et al. 1998 Huffcutt et al. 2001 Hui, Lam & Law 2000 Johnson 2001 Judge et al. 1999 Kaufman, Stamper & Tesluk 2001 Kidder 2002 Kinicki, Lockwood & Hom 1990 Koh, Steers & Terborg 1995 Konovsky & Organ 1996 Lam, Hui & Law 1999 Latham & Skarlicki 1995 Lee & Allen 2002 LePine & Van Dyne 2001 LePine, Colquitt & Erez 2000 Love et al. 1994 MacKenzie, Podsakoff & Fetter 1991 MacKenzie, Podsakoff & Rich 2001 MacKeznzie, Podsakoff, Paine 1999 Mael & Ashforth 1995 McDaniel 1989 McHenry et al. 1990 McManus & Kelly 1999 McNeely & Meglino 1994 Mengue 2000 Miller, Griffin & Hart 1999 Moorman & Blakely 1995 Moorman 1991 Moorman 1993 Moorman, Niehoff & Organ 1993 Morrison 1994 Motowidlo & Van Scotter 1994 Motowidlo et al. 1992

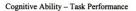
Mount et al. 1998 Mount, Witt & Barrick 2000 Mumford et al. 1996 Nathan & Alexander 1988 Nathan & Tippins 1990 Neuman & Kickul 1998 Neuman & Wright 1999 Niehoff & Moorman 1993 Nikolaou & Robertson 2002 O'Connell et al. 2001 O'Connell et al. 2001 Organ & Konovsky 1989 Piedmont & Weinstein 1994 Ployhart, Lim & Chan 2001 Ployhart et al. 2003 Podsakoff & Mackenzie 1994 Podsakoff, MacKenzie & Bommer 1996 Podsakoff, MacKenzie & Fetter 1993 Podsakoff et al. 1990 Pulakos & Schmitt 1996 Pulakos, Borman & Hough 1988 Randall et al. 1999 Ree, Caretta & Teachout 1995 Ree, Earles & Teachout 1994 Rioux & Penner 2001 Russell et al. 1990 Rvan, Ployhart & Friedel 1998 Sackett, Gruys & Ellingson 1998 Schmidt & Rader 1999 Schmidt et al. 1988 Schmitt & Ryan 1993 Schnake, Dumler & Cochran 1993 Scullen, Mount & Judge 2003 Shore & Wayne 1993 Shore, Barksdale & Shore 1995 Shore et al. 2000 Stewart 1996 Stokes & Searcy 1999 Tansky 1993 Tepper, Lockhart & Hoobler 2001 Tompson & Werner 1997 Turnipseed 2002 Turnipseed 2003

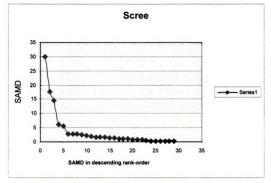
Appendix F

(continued)

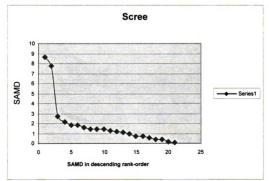
Turnley et al. 2003 Van Dyne, Graham & Dienesch 1994 Van Scotter & Motowidlo 1996 VanYperen, van den Berg & Willering 1999 Villanova et al. 1994 Williams & Anderson 1991 Witt, Burke, Barrick, & Mount 2002 Mount, Barrick & Stewart 1998 Mount, Witt & Barrick 2000 Barrick & Mount 1996 Yoon & Suh 2003

Appendix G Scree Plots for Outlier Analyses



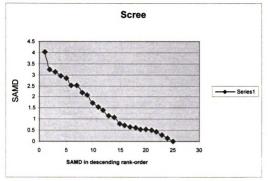


Conscientiousness - Task Performance

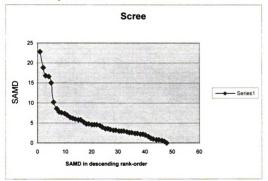




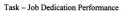


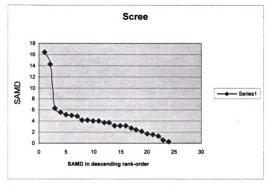


Task - Citizenship Performance

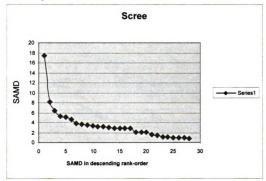






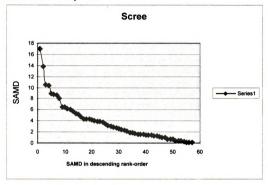


Task - Interpersonal Performance

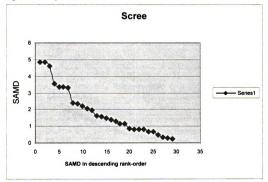


Appendix G (continued)



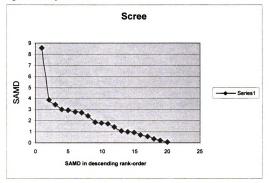


Cognitive Ability - Conscientiousness

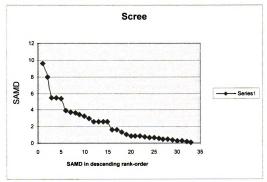






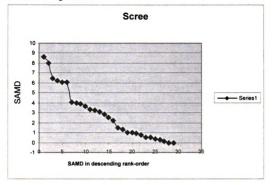


Extraversion - Conscientiousness

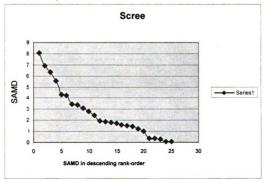






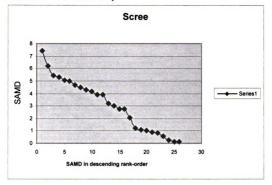


Extraversion - Openness

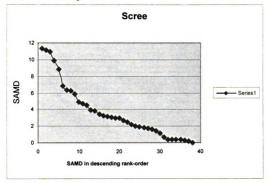






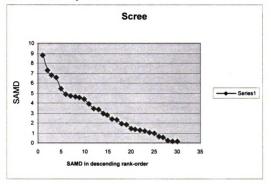


Conscientiousness - Agreeableness

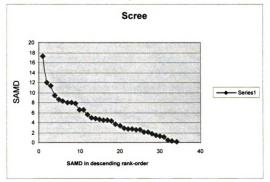


Appendix G (continued)

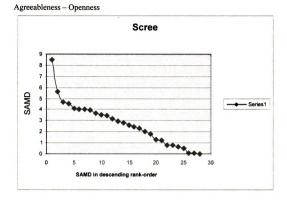




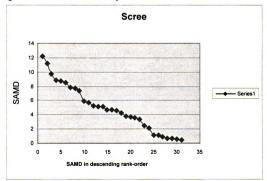
Conscientiousness - Emotional Stability



Appendix G (continued)

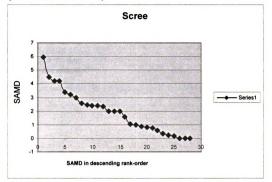


Agreeableness - Emotional Stability









___ =

Appendix H Job Complexity Codes

Study	Complexity
•	
Allworth & Hesketh 1999	0
Hedge & Teachout 1992	0
Johnson 2001	1
Pulakos & Schmitt 1996	1
Pulakos, Borman & Hough 1988	0
Ree, Caretta & Teachout 1995	1
Ree, Earles & Teachout 1994	1

Note. 0 = low complexity job, 1 = high complexity, 2 = indeterminate

Appendix I Biodata Studies For Which Raters Assigned Construct Codes

k

Study	<u>N</u>	Cognitive (%)	Personality (%)
Allworth & Hesketh 1999	169	3	57
Borman et al. 1991	4,362	0	65
Bosshardt et al. 1992	357	18	53
Dalessio & Silverhart 1994	577	1	25
Hough et la. 1990	7,666	0	58
McDaniel 1990	9,336	7	27
McHenry et al. 1990	4,039	0	56
McManus & Kelly 1999	116	2	37
Mumford et al. 1996	117	5	36
O'Connell et al. 2001	94	8	20
Pulakos & Schmitt 1996	461	8	49
Russell et al. 1990	273	13	36
Stokes & Searcy 1999	933	8	40

sampres.
c nataline
OI LCSI
contro 1

		-	2	e	4	S	9	7	∞	6	10	11
Ξ	(1) Cognitive Ability											
(5)		.03										
(3)	Conscientiousness	60.	.29									
(4	Agreeableness	.03	.29	.51								
(5)		.21	.34	60.	.13							
9		.19	.38	.67	.56	.18						
6	Biodata	.27	.31	.37	.37	.33	.45					
(8)	Task Performance	.21	.05	60.	.05	.02	.12	.16				
6	Citizenship Perf.	.29	<u>.</u>	.20	.17	.05	.11	.30	.50			
(10)	Job Dedication	60.	00.	.17	.15	90.	.06	.31	.37	•		
(11)	Interpersonal Facil.	.04	.03	.13	.20	.01	90.	.48	.33	ı	.72	
(12)	(12) Overall Perf.	.11	.01	.20	90.	.11	.07	.26	.41	.65	.61	.59

Note. Values are correlations corrected for unreliability. Results are not provided for unrestricted samples because there were only 25 studies. The results here are shown to suggest that some correlations are attenuated due to range restriction.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- *Ackerman, P. L. & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology*, 78, 413-432.
- Algera, J. A., Jansen, P. G., Roe, R. A., & Vijn, P. (1984). Validity generalization: Some critical remarks on the Schmidt-Hunter procedure. *Journal of Occupational Psychology*, 57, 197-210.
- *Allen, T. D. & Rush, M. C. (1998). The effects of organizational citizenship behavior on performance judgments: a field study and a laboratory experiment. *Journal of Applied Psychology*, 83, 247-260.
- *Allworth, E., & Hesketh, B. (1999). Construct-oriented biodata: Capturing changerelated and contextually relevant future performance. *International Journal of Selection & Assessment Background*, 7, 97-111.
- *Antonioni, D. & Park, H. (2001). The effects of personality similarity on peer ratings of contextual work behaviors. *Personnel Psychology*, 54, 331-360.
- Ashworth, S. D., Osburn, H. G., Callender, J. C., & Boyle, K. A. (1992). The effects of unrepresented studies on the robustness of validity generalization results. *Personnel Psychology*, 45, 341-361.
- Austin, J. T. & Villanova, P. (1992). The criterion problem: 1917-1992. Journal of Applied Psychology, 77, 836-874.
- Avolio, B. J. & Waldman, D. A. (1990). An examination of age and cognitive test performance across job complexity and occupational types. *Journal of Applied Psychology*, 75, 43-50.
- *Barbuto, J. E., Jr., Brown, L. L., Wheeler, D. W., & Wilhite, M. S. (2003). Motivation, altruism and generalized compliance: A field study of organizational citizenship behaviors. *Psychological Reports*, *92*, 498-502.
- *Barksdale, K., & Werner, J. M. (2001). Managerial ratings of in-role behaviors, organizational citizenship behaviors and overall performance: Testing different models of their relationship. *Journal of Business Research*, 51, 145-155.
- Barnard C. I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.

- Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44, 1-26.
- *Barrick, M. R. & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the big five personality dimensions and job performance. *Journal of Applied Psychology*, 78, 111-118.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where to we go next? *Personality and Performance*, 9, 9-30.
- *Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effect of goal setting. *Journal of Applied Psychology*, 78, 715-722.
- *Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, 87, 43-51.
- *Beaty, J. C. Jr., Cleveland, J. N., & Murphy, K. R. (2001). The relation between personality and contextual performance in "strong" versus "weak" situations, *Human Performance*, 14, 125-148.
- Becker, B. J. (1992). Using results from replicated studies to estimate linear models. Journal of Educational Statistics, 17, 341-362.
- Becker, B. J. (1996). The generalizability of empirical research results. In C. P. Benbow & D. Lubinksi (Eds.) *Psychometric and social issues: Intellectual talent*. (pp. 362-383). Baltimore, MD: The Johns Hopkins University Press.
- Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & S. D. Brown (Eds.) Handbook of applied multivariate statistics and mathematical modeling. (pp. 499-525). San Diego: Academic Press.
- *Becker, T. E., & Vance, R. J. (1993). Construct validity of three types of organizational citizenship behavior: An illustration of the direct product model with refinements. *Journal of Management, 19*, 663-682.
- *Bell, B. S. & Kozlowski, S. W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology*, 87, 497-505.
- *Bell, S. J., & Menguc, B. (2002). The employee-organization relationship, organizational citizenship behaviors, and superior service quality. *Journal of Retailing*, 78, 131-146.

- Bernardin, H.J. & Beatty, R.W. (1984). Performance appraisal: Assessing human behavior at work. Boston:Kent.
- Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1997). Theoretical bases for team self-corrections: Fostering shared mental models. In M. M. Beyerlein & D. A. Johnson (Eds.), Advances in interdisciplinary studies of work teams, Vol. 4 (pp. 249-279). US: Elsevier Science/JAI Press.
- Bliese, P. D. (2002). Multilevel random coefficient modeling in organizational research: Examples using SAS and S-Plus. In F. Drasgow & N. Schmitt (Eds.), Measuring and analyzing behavior in organizations: Advances in measurement and Data Analysis (pp. 401–445). San Francisco, CA: Jossey-Bass.
- Bobko, P., Roth, P., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, 52, 561-589.
- Bommer, W. H., Johnson, J., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology*, 48, 587-606.
- Borman, W. C. (1982). Validity of behavioral assessment for predicting military recruiter performance, *Journal of Applied Psychology*, 67, 3-9.
- Borman, W. C. & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations*, (pp. 71-98). San Francisco: Jossey-Bass.
- Borman, W. C. & Motowidlo, S. J. (1997). Task performance and contextual performance: the meaning for personnel selection research. *Human Performance*, 10, 99-109.
- Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment* 9, 52-69.
- *Borman, W. C., White, L. A. & Dorsey, D. W. (1995). Effects of rate task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology*, 80, 168-177.
- *Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology*, 76, 863-872.
- *Bosshardt, M. J., Carter, G. W., Gialluca, K. A., Dunnette, M. D., & Ashworth, S. D. (1992). Predictive validation of an insurance agent support person selection battery. *Journal of Business & Psychology Test validity yearbook: I*, 7, 213-224.

- *Botwin, M. D. & Buss, D. M. (1989). Structure of act-report data: Is the five-factor model of personality recaptured? *Journal of Personality and Social Psychology*, 56, 988-1001.
- *Boudreau, J. W., Boswell, W. R., Judge, T. A., & Bretz, R. D. Jr. (2001). Personality and cognitive ability as predictors of job search among employed managers. *Personnel Psychology*, 54, 25-50.
- Brief, A. P. & Motowidlo, S. J. (1986). Prosocial organizational behaviors. Academy of Management Review, 11, 710-725.
- *Brown, S. H., Stout, J. D., Dalessio, A. T., & Crosby, M. M. (1988). Stability of validity indices through test score ranges. *Journal of Applied Psychology*, 73, 736-742.
- *Burroughs, W. A. & White, L. L. (1996). Predicting sales performance. Journal of Business and Psychology, 11, 73-84.
- *Caldwell, D. F. & Burger, J. M. (1998). Personality characteristics of job applicants and success in screening interviews. *Personnel Psychology*, 53, 119-136.
- *Caliguiri, P. M. (2000). The big five personality characteristics as predictors of expatriate's desire to terminate the assignment and supervisor rated-performance. *Personnel Psychology*, 53, 67-88.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology, In M. D. Dunnette & L. M. Hough (Eds.), *Handbook* of industrial and organizational psychology (2nd ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E. III, & Weick, K. R., Jr. (1970). Managerial behavior, performance, and effectiveness. New York: McGraw-Hill.
- Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of job performance variability. In K. R. Murphy (Ed.) *Individual differences and behavior in organizations* (pp. 258-299). San Francisco, CA: Jossey-Bass Publishers.
- Campbell, J. P, McCloy, R. A., Oppler, S. H. & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in* organizations, (pp. 2-69). San Francisco: Jossey-Bass.
- Campbell, J. T., Prien, E. P., & Brailey, L. G. (1960). Predicting performance evaluations. *Personnel Psychology*, 13, 435-440.

- *Campion, M. A., Campion, J. E. & Hudson, J. P., Jr.(1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, *79*, 998-1002.
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50, 655-702.
- *Campion, M. A., Pursell, E. D. & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology*, 41, 25-42.
- Carr, J. Z., Schmidt, A. M., Ford, J. K., & DeShon, R. P. (2003). Climate perceptions matter: A meta-analytic path analysis relating molar climate, cognitive and affective states, and individual level work outcomes. *Journal of Applied Psychology*, 88, 605-619.
- Cascio, W. F. (1995). Whither industrial and organizational psychology in a changing world of work? *American Psychologist*, 50, 928-939.
- *Charbonneau, D., & Nicol, A. A. M. (2002). Emotional intelligence and prosocial behaviors in adolescents. *Psychological Reports*, 90, 361-370.
- *Chen, Z. X., & Francesco, A. M. (2003). The relationship between the three components of commitment and employee performance in China. *Journal of Vocational Behavior*, 62, 490-510.
- Cleveland, J. N., Murphy, K. R. & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, 74, 130-135.
- Cohen, J. (1977). Statistical power analysis for the behavioral sciences (Revised edition). New York: Academic Press.
- Coleman, V. I. & Borman, W. C. (2000). Investigating the underlying structure of the citizenship performance domain. *Human Resource Management Review*, 10, 25-44.
- *Collins, J. M. & Gleaves, D. H. (1998). Race, job applicants, and the five-factor model of personality: Implications for black psychology, industrial/organizational psychology, and the five-factor theory. *Journal of Applied Psychology*, 83, 531-544.
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678-707.

- Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin, 88*, 322-328.
- Conway, J. M. (1996). Additional construct validity evidence for the task/contextual performance distinction. *Human Performance*, 9, 309-329.
- *Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs, *Journal of Applied Psychology*, 84, 3-13.
- Cooper, H. (1998). Synthesizing research, Sage Publications, Thousand Oaks, CA.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- *Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G., & Smith, R. G. (1992). The "big five" personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology*, 45, 119-140.
- *Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores, *Personnel Psychology*, 53, 325-351.
- *Costa, P. T. & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, 55, 258-265.
- *Crant, J. M. (1995). The proactive personality scale and objective job performance among real estate agents. *Journal of Applied Psychology*, 80, 532-537.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cronbach, L. J. & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418.
- *Cropanzano, R., Rupp, D. E.,& Bryne, Z. S. (2003). The relationship of emotional exhaustion to work attitudes, job performance, and organizational citizenship behaviors. *Journal of Applied Psychology*, 88, 160-169.
- *Dalessio, A. T., & Silverhart, T. A. (1994). Combining biodata test and interview information: Predicting decisions and performance criteria. *Personnel Psychology*, 47, 303-315.
- Darlington, R. B. (1990). Regression and linear models. New York: McGraw-Hill Publishing Co.

- *Day, D. V. & Silverman, S. B. (1989). Personality and job performance: evidence of incremental validity. *Personnel Psychology*, 42, 25-36.
- De Corte, W. (1999). Weighing job performance predictors to both maximize quality of the selected workforce and control the level of adverse impact. *Journal of Applied Psychology*, 84, 695-702.
- *De Fruyt, F. & Mervielde, I. (1999). RIASEC types and big five traits as predictors of employment status and nature of employment. *Personnel Psychology*, 52, 701-727.
- *Deadrick, D. L. & Madigan, R. M. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology*, 43, 717-744.
- Dickinson, T. L. & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick & E. Salas (Eds.), *Team performance* assessment and measurement: Theory, methods, and applications (pp. 19-43). Mahwah, NJ: Lawrence Erlbaum Associates.
- *Douthitt, S. S., Eby, L. T., & Simon, S. A. (1999). Diversity of life experiences: The development and validation of a biographical measure of receptiveness to dissimilar others. *International Journal of Selection & Assessment Background*, 7, 112-125.
- *Farh, J.-l., Podsakoff, P. M., & Organ, D. W. (1990). Accounting for Organizational Citizenship Behavior: Leader fairness and task scope versus satisfaction. *Journal* of Management, 16, 705-721.
- *Farh, J.-l., Werbel, J. D. & Bedeian, A. G. (1988). An empirical investigation of selfappraisal-based performance evaluation. *Personnel Psychology*, 41, 141-156.
- *Ferris, G. R., Witt, L. A. & Hochwarter, W. A. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology*, 86, 1075-1082.
- *Findley, H. M., Giles, W. F. & Mossholder, K. W. (2000). Performance appraisal process and system facets: Relationships with contextual performance. *Journal of Applied Psychology*, 85, 634-640.
- Ford, J. K., Kraiger, K. & Schechtman, S. L. (1986). Study of race effects in objective indices and subjective evaluations of performance: a meta-analysis of performance criteria. *Psychological Bulletin*, 99, 330-337.
- *Gellatly, I. R. (1996). Conscientiousness and task performance: Test of a cognitive process model. *Journal of Applied Psychology*, 81, 474-482.

- *Gellatly, I. R., & Irving, P. G. (2001). Personality, autonomy, and contextual performance. *Human Performance*, 14, 231-245.
- George, J. M. & Brief, A. P. (1992). Feeling good-doing good: a conceptual analysis of the mood at work-organizational spontaneity relationship. *Psychological Bulletin*, *112*, 310-329.
- *George, J. M. (1991). State or trait: Effects of positive mood on prosocial behaviors at work. *Journal of Applied Psychology*, 76, 299-307.
- Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L.
 V. Hedges (Eds.). *The handbook of research synthesis* (pp. 339-355). New York: Russell Sage Foundation.
- *Goffin, R. D., Rothstein, M. G. & Johnston, N. G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology*, 81, 746-756.
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. Intelligence, 24, 79-132.
- Guion, R. M. & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 135-164.
- Guion, R. M. (1998). Assessment, measurement, and prediction for personnel decisions. Lawrence Erlbaum Associates, Mahwah: New Jersey.
- *Hansen, C. P. (1989). A causal model of the relationship among accidents, biodata, personality, and cognitive factors. *Journal of Applied Psychology*, 74, 81-90.
- Hattie, J. A., & Hansford, B.C. (1984). Meta-analysis: A reflection on problems. Australian Journal of Education, 36, 239-254.
- *Hattrup, K., O'Connell, M. S., & Wingate, P. H. (1998). Prediction of multidimensional criteria: distinguishing task and contextual performance. *Human Performance*, 11, 305-319.
- Hattrup, K., Rock, J. & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology*, 82, 656-664.
- *Haworth, C. L., & Levy, P. E. (2001). The importance of instrumentality beliefs in the prediction of organizational citizenship behaviors. *Journal of Vocational Behavior*, 59, 64-75.
- *Hedge, J. W. & Teachout, M. S. (1992). An interview approach to work sample criterion measurement. *Journal of Applied Psychology*, 77, 453-461.

- Hedges, L. V. & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6, 203-217.
- *Hochwarter, W. A., Witt, L. A. & Kacmar, K. M. (2000). Perceptions of organizational politics as a moderator of the relationship between conscientiousness and job performance. *Journal of Applied Psychology*, 85, 472-478.
- *Hogan, J., Hogan, R. & Gregory, S. (1992). Validation of a sales representative selection inventory. *Journal of Business and Psychology*, 7, 161-171.
- *Hogan, J., Rybicki, S. L., Motowidlo, S. J., & Borman, W. C. (1998). Relations between contextual performance, personality, and occupational advancement. *Human Performance*, 11, 189-207.
- Hom, P. W., Caranika-Walker, F., Prussia, G. E., & Griffeth, R. W. (1992). A metaanalytical structural equations analysis of a model of employee turnover. *Journal* of Applied Psychology, 77, 890-909.
- *Hough, L. M. (1992). The "big five" personality variables construct confusion: Description versus prediction. *Human Performance*, 5, 139-155.
- *Hough, L. M., Eaton, N. K., Dunnette, M. D., & Kamp, J. D. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 5, 581-595.
- Hough, L. M. & Oswald, F. L. (2000). Personnel selection: looking toward the future remembering the past. *Annual Review of Psychology*, 51, 631-664.
- Huffcutt, A. I. & Arthur, W. A. Jr. (1995). Development of a new outlier statistic for meta-analytic data. Journal of Applied Psychology, 80, 327-334.
- *Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897-913.
- Huffcutt, A. I. & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, 83, 179-189.
- Huffcutt, A., Roth, P. & McDaniel, M. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: moderating characteristics and implications of incremental validity. *Journal of Applied Psychology*, 81, 459-473.
- *Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Degroot, T. G., & Jones, C. (2001). Comparison of situational behavior description interview questions for higherlevel positions. *Personnel Psychology*, 54, 619-644.

- *Hui, C., Lam, S. S. & Law, K. K. S. (2000). Instrumental values of organizational citizenship behavior for promotion: A field quasi-experiment. *Journal of Applied Psychology*, 85, 822-828.
- Hunt, S. T. (2002). On the virtues of staying 'inside the box': Does organizational citizenship behavior detract from performance in Taylorist jobs? *International Journal of Selection and Assessment*, 10, 152-159
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings* (1st ed). London: Sage Publications.
- Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd Edition). Thousand Oaks, CA: Sage Publications.
- Hurtz, G. M. & Donovan, J. J. (2000). Personality and job performance: the Big Five revisited. *Journal of Applied Psychology*, 85, 869-879.
- Ilgen, D. R., & Hollenbeck, J. R. (1991). Job design and roles. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed., Vol. 2). Palo Alto, CA: Consulting Psychologists Press.
- Jensen, A. R. (1998). The g factor: The science of mental ability. Westport, CT: Praeger.
- John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.) Handbook of personality: Theory and research. New York: Guilford.
- *Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology*, *86*, 984-996.
- *Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621-652.
- Kalaian, H. A. & Raudenbush, S. W. (1996). A multivariate linear model for metaanalysis. *Psychological Methods*, 1, 227-235.
- Katz, D. (1964). The motivational basis of organizational behavior. *Behavioral Science*, 9, 131-146.
- *Kaufman, J. D., Stamper, C. L., & Tesluk, P. E. (2001). Do supportive organizations make for good corporate citizens? *Journal of Managerial Issues*, 13, 436-449.

- Kelloway, E. K., Loughlin, C., Barling, J. & Nault, A. (2002). Self-reported counterproductive behaviors and organizational citizenship behaviors: separate but related constructs. *International Journal of Selection and Assessment*, 10, 143-151.
- Kenny, D. A. & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin*, 119, 138-148.
- *Kidder, D. L. (2002). The influence of gender on the performance of organizational citizenship behaviors. *Journal of Management, 28*, 629-648.
- *Kinicki, A. J., Lockwood, C. A., Hom, P. W., & Griffeth, R. W. (1990). Interviewer predictions of applicant qualifications and interviewer validity: Aggregate and individual analyses. *Journal of Applied Psychology*, *75*, 477-486.
- Kline, R. B. (1998). Principles and practice of structural equation modeling. New York: The Guilford Press.
- *Koh, W. L., Steers, R. M., & Terborg, J. R. (1995). The effects of transformation leadership on teacher attitudes and student performance in Singapore. *Journal of Organizational Behavior*, 16, 319-333.
- *Konovsky, M. A., & Organ, D. W. (1996). Dispositional and contextual determinants of organizational citizenship behavior. *Journal of Organizational Behavior*, 17, 253-266.
- Koslowsky, M. & Sagie, A. (1993). On the efficacy of credibility intervals as indicators of moderator effects in metaanalytic research. *Journal of Organizational Behavior*, 14, 695-699.
- Koslowsky, M. & Sagie, A. (1994). Components of artifactual variance in meta-analytic research. *Personnel Psychology*, 47, 561-574.
- Kozlowski, S. W. J., S. M. Gully, et al. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. In D. R. Ilgen and E. D. Pulakos (Eds.), *The changing nature of performance: Implications for staffing, motivation, and development* (pp. 240-292). San Francisco: Josey-Bass.
- Kravitz, D. A. & Balzer, W. K. (1992). Context effects in performance appraisal: A methodological critique and empirical study. *Journal of Applied Psychology*, 77, 24-31.
- Lacayo, R. & Ripley, A. (2002, December 30/2003, January 6). Persons of the year. *TIME*, 30-33.

- *Lam, S. S. K., Hui, C. & Law, K. S. (1999). Organizational citizenship behavior: Comparing perspectives of supervisors and subordinates across four international samples. *Journal of Applied Psychology*, *84*, 594-601.
- Lance, C. E. & Bennett, W. (2000). Replication and extension models of supervisory job performance ratings. *Human Performance*, 13, 139, 158.
- Landis, J. & Koch, G. G. (1977). The measurement of observer agreement for categorical data, *Biometrics*, 33, 159-174.
- *Latham, G. P. & Skarlicki, D. P. (1995). Criterion-related validity of the situational and patterned behavior description interviews with organizational citizenship behavior, *Human Performance*, 8, 67-80.
- Lawler, E. E., Mohrman, S. A., & Ledford, G. E. (1995). Creating high performance organizations: Practices and results of employee involvement and total quality management in Fortune 1000 companies. San Francisco: Jossey-Bass.
- *Lee, K. & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: the role of affect and cognitions. *Journal of Applied Psychology*, 87, 131-142.
- *LePine, J. A. & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: evidence of differential relationships with big five personality characteristics and cognitive ability. *Journal of Applied Psychology*, 86, 326-336.
- *LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology*, 53, 563-593.
- LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: a critical review and meta-analysis. *Journal of Applied Psychology*, 87, 52-65.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Locke, E. A. (1982). The ideas of Frederick W. Taylor: An evaluation. Academy of Management Review, 7, 14-24.
- *Love, K. G., Bishop, R. C., Heinisch, D. A., & Montei, M. S. (1994). Selection across two cultures: Adapting the selection of American assemblers to meet Japanese job performance demands. *Personnel Psychology*, 47, 837-846.

- Lovell, S. E., Kahn, A. S., Anton, J., Davidson, A., Dowling, E., Post, D., et al. (1999). Does gender affect the link between organizational citizenship behavior and performance evaluation? *Sex Roles*, 41, 469-478.
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points." *Annual Review of Psychology*, 51, 405-444.
- *MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons' performance. Organizational Behavior & Human Decision Processes, 50, 123-150.
- *MacKenzie, S. B., Podsakoff, P. M., & Paine, J. B. (1999). Do citizenship behaviors matter more for managers than for salespeople? *Journal of the Academy of Marketing Science*, 27, 396-410.
- *MacKenzie, S. B., Podsakoff, P. M., & Rich, G. A. (2001). Transformational and transactional leadership and salesperson performance. *Journal of the Academy of Marketing Science, 29*, 115-134.
- *Mael, F. A., & Ashforth, B. E. (1995). Loyal from day one: Biodata, organizational identification, and turnover among newcomers. *Personnel Psychology*, 48, 309-333.
- Mathieu, J. E. & Zajac, D. M. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin*, 108, 172-194.
- Mayberry, P. W. & Carey, N. B. (1997). The effect of aptitude and experience on mechanical job performance, *Educational and Psychological Measurement*, 57, 131-149.
- McCloy, R. A., Campbell, J. P. & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology*, 79, 493-505.
- *McDaniel, M. A. (1989). Biographical constructs for predicting employee suitability. Journal of Applied Psychology, 74, 964-970.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal* of Applied Psychology, 79, 599-616.
- *McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: the relationship between predictor and criterion domains, *Personnel Psychology*, 43, 335-354.

- *McManus, M. A. & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry, *Personnel Psychology*, 52, 137-148.
- *McNeely, B. L. & Meglino, B. M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. *Journal of Applied Psychology*, 79, 836-844.
- *Menguc, B. (2000). An empirical investigation of a social exchange model of organizational citizenship behaviors across two sales situations: A Turkish case. *Journal of Personal Selling & Sales Management, 20*, 205-214.
- Miles, D. E., Borman, W. E., Spector, P. E., & Fox, S. (2002). Building an integrative model of extra role work behaviors: a comparison of counterproductive work behavior with organizational citizenship behavior. *International Journal of Selection and Assessment*, 10, 51-57.
- *Miller, R. L., Griffin, M. A. & Hart, P. M. (1999). Personality and organizational health: The role of conscientiousness. *Work and Stress*, 13, 7-19.
- Mitchell, T. W. (1994). The utility of biodata. In G. S. Stokes, M. D. Mumford, & M. A. Owens (Eds.), *Biodata Handbook* (pp. 485-516). Palo Alto, CA: CPP Books.
- *Moorman, R. H. (1991). Relationship between organizational justice and organizational citizenship behaviors: Do fairness perceptions influence employee citizenship? Journal of Applied Psychology, 76, 845-855.
- *Moorman, R. H. (1993). The influence of cognitive and affective based job satisfaction measures on the relationship between satisfaction and organizational citizenship behavior. *Human Relations, 46*, 759-776.
- *Moorman, R. H. & Blakely, G. L. (1995). Individualism^collectivism as an individual difference predictor of organizational citizenship behavior. *Journal of Organizational Behavior*, 16, 127-142.
- *Moorman, R. H., Niehoff, B. P. & Organ, D. W. (1993). Treating employees fairly and organizational citizenship behavior: Sorting the effects of job satisfaction, organizational commitment, and procedural justice. *Employee Responsibilities and Rights Journal*, 6, 209-225.
- *Morrison, E. W. (1994). Role definitions and organizational citizenship behavior: The importance of the employee's perspective. Academy of Management Journal, 37, 1543-1567.
- Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10, 71-83.

- *Motowidlo, S. J., Carter, G. W., Dunette, M. D., Tippins, N., Werner, S. Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal* of Applied Psychology, 77, 571-587.
- *Motowidlo, S. J. & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475-480.
- *Mount, M. K., Judge, T. A., Scullen, S. E., Systma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, 51, 557-576.
- *Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology*, 53, 299-323.
- Muchinsky, P. (1997) *Psychology applied to work (Fifth Edition)*. Pacific Grove, CA: Brooks/Cole Publishing.
- Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement*, 11, 1-31.
- *Mumford, M. D., Costanza, D. P., Connelly, M. S., & Johnson, J. F. (1996). Item generation procedures and background data scales: Implications for construct and criterion-related validity. *Personnel Psychology*, 49, 361-398.
- Murphy, K. R. (1989). Dimensions of job performance. In R. Dillon & J. Pellingrino (Eds.), *Testing: Applied and theoretical perspectives* (pp. 218-247). New York: Praeger.
- Murphy, K. R. (1997). Meta-analysis and validity generalization. In N. Anderson & P. Herriot (Eds.) *International Handbook of Selection and Assessment*, (pp. 323-342). Cinchester: John Wiley & Sons Ltd.
- Murphy, K. R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance*, 15, 173-186.
- Murphy, K. R. & Cleveland, J. N. (1995). Understanding performance appraisal: Social, organizational, and goal-based perspectives. Thousand Oaks, CA: Sage.
- Murphy, K. R. & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, 53, 873-900.

- Murphy, K. R. & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology*, 50, 823-854.
- *Nathan, B. R. & Alexander, R. A. (1988). A comparison of criteria for test validation: a meta-analytic investigation, *Personnel Psychology*, 41, 517-535.
- *Nathan, B. R. & Tippins, N. (1990). The consequences of halo "error" in performance ratings: A field study of the moderating effect of halo on test validation results. *Journal of Applied Psychology*, 75, 290-296.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996).
 Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.
- *Neuman, G. A. & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology*, 84, 376-389.
- *Neuman, G. A., & Kickul, J. R. (1998). Organizational citizenship behaviors: Achievement orientation and personality. *Journal of Business & Psychology*, 13, 263-279.
- Nickels, B. J. (1994). The nature of biodata. In G. S. Stokes, M. D. Mumford, & M. A. Owens (Eds.), *Biodata Handbook* (pp. 1-16). Palo Alto, CA: CPP Books.
- *Niehoff, B. P. & Moorman, R. H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. The *Academy of Management Journal*, 36, 527-556.
- *Nikolaou, I., & Robertson, I. T. (2001). The Five-Factor model of personality and work behaviour in Greece. European Journal of Work & Organizational Psychology, 10, 161-186.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw Hill, Inc.
- *O'Connell, M. S., Doverspike, D., Norris-Watts, C., & Hattrup, K. (2001). Predictors of organizational citizenship behavior among Mexican retail salespeople. International Journal of Organizational Analysis, 9, 272-280.
- *O'Connell, M. S., Hattrup, K., Doverspike, D., & Cober, A. (2002). The validity of "mini" simulations for Mexican retail salespeople. *Journal of Business & Psychology*, 16, 593-600.
- Organ, D. W. (1988). Organizational citizenship behavior: The good soldier syndrome. Lexington, MA: Lexington Books.

- Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. Human Performance, 10, 85-97.
- *Organ, D. W. & Konovsky, M. (1989). Cognitive versus affective determinants of organizational citizenship behavior. *Journal of Applied Psychology*, 74, 153-164.
- Organ, D. W. & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, 48, 775-802.
- Orwin, R. G. & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychological Bulletin*, 97, 134-147.
- Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. Human Performance, 15, 161-172.
- Paese, P. W. & Switzer, F. S. III (1988). Validity generalization and hypothetical reliability distributions: A test of the Schmidt-Hunter Procedure. *Journal of Applied Psychology*, 73, 267-274.
- Penner, L. A., Midili, A. R. & Kegelmeyer, J. (1997). Beyond job attitudes: A personality and social psychology perspective on the causes of organizational citizenship behavior. *Human Performance*, 10, 111-131.
- Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. L., Houston, J. S., Toquam, J. L., & Wing, H. (1990). Project A: Specification of the predictor domain and development of new selection/classification tests, *Personnel Psychology*, 43, 247-276.
- *Phillips, J. M. & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in self-efficacy and goal-setting process. *Journal of Applied Psychology*, 82, 792-802.
- *Piedmont, R. L. & Weinstein, H. P. (1994). Predicting supervisor ratings of job performance using the NEO personality inventory. *The Journal of Psychology*, *128*, 255-265.
- *Ployhart, R. E., Lim, B. & Chan, K. (2001). Exploring relations between typical and maximum performance ratings and the five factor model of personality. *Personnel Psychology*, 54, 809-843.
- *Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paperand-pencil testing of applicants in a proctored setting: Are personality, biodata and situational judgment tests comparable? *Personnel Psychology*, 56, 733-752.

- *Podsakoff, P. M. & MacKenzie, S. B. (1994). Organizational citizenship behaviors and sales unit effectiveness. *Journal of Marketing Research*, 31, 351-363.
- *Podsakoff, P. M., MacKenzie, S. B., & Bommer, W. H. (1996). Transformational leader behaviors and substitutes for leadership as determinants of employee satisfaction, commitment, trust, and organizational citizenship behaviors. *Journal of Management*, 22, 259-298.
- *Podsakoff, P. M., MacKenzie, S. B., & Fetter, R. (1993). Substitutes for leadership and the management of professionals. *The Leadership Quarterly*, 4, 1-44.
- *Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H. & Fetter, R. (1993). Transformational leader behaviors and their effects on followers' trust in leader. *The Leadership Quarterly*, 1, 107-142.
- Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, 26, 513-563.
- Puffer, S. M. (1987). Prosocial behavior, noncompliant behavior, and work performance among commission salespeople. *Journal of Applied Psychology*, 72, 615-621.
- *Pulakos, E. D. & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance*, 9, 241-258.
- *Pulakos, E. D., Borman, W. C. & Hough, L. M. (1988). Test validation for scientific understanding: Two demonstrations of an approach to studying predictor-criterion linkages. *Personnel Psychology*, 41, 703-716.
- Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology*, 74, 770-780.
- Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology*, 76, 432-446.
- Raju, N. S., Pappas, S. & Williams, C. P. (1989). An empirical monte carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. *Journal of Applied Psychology*, 74, 901-911.
- *Randall, M. L., Cropanzano, R., Bormann, C. A., & Birjulin, A. (1999). Organizational politics and organizational support as predictors of work attitudes, job performance, and organizational citizenship behavior. *Journal of Organizational Behavior*, 20, 159-174.

- Raudenbush, S. W., Becker, B. J. & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, 103, 111-120.
- Ree, M. J. & Carretta, T. R. (2002). g2K. Human Performance, 15, 3-23.
- *Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology*, 80, 721-730.
- *Ree, M. J., Earles, J. A. & Teachout, M. S. (1994). Predicting job performance: Not much more than g. Journal of Applied Psychology, 79, 518-524.
- *Rioux, S. M. & Penner, L. A. (2001). The causes of organizational citizenship behavior: A motivational analysis. *Journal of Applied Psychology*, 86, 1306-1314.
- Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology*, 75, 322-327.
- Rothstein, H. R., McDaniel, M. A., & Borenstein, M. (2002). Meta-analysis: A Review of Quantitative Cumulation Methods. In F. Drasgow & N. Schmitt (Eds.) *Measuring and analyzing behavior in organizations* (pp. 534-570). San Francisco: Jossey-Bass, Inc.
- Rotton, J., Foos, P. W., Van Meek, L., & Levitt, M. (1995). Publication practices and the file drawer problem: A survey of published authors. *Journal of Social Behavior* and Personality, 10, 1-13.
- Rotundo, M. & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy capturing approach. *Journal of Applied Psychology*, 87, 66-80.
- Russell, C. J. & Gilliland, S. W. (1995). Why meta-analysis doesn't tell us what the data really mean: distinguishing between moderator effects and moderator processes. *Journal of Management*, 21, 813-831.
- *Russell, C. J., Mattson, J., Devlin, S. E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology*, 75, 569-580.
- *Ryan, A. M., Ployhart, R. E. & Friedel, L. A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology*, 83, 298-302.
- *Sackett, P. R., Gruys, M. L. & Ellingson, J. E. (1998). Ability-personality interactions when predicting job performance. *Journal of Applied Psychology*, 83, 545-556.

- Sackett, P. R., Schmitt, N., Ellingson, J. E. & Kabin, M. B. (2001). High-stakes testing employment, credentialing, and higher education: Prospects in a post-affirmativeaction world. *American Psychologist*, 56, 302-318.
- Sagie, A. & Koslowsky, M. (1993). Detecting moderators with meta-analysis: An evaluation and comparison of techniques. *Personnel Psychology*, 46, 629-640.
- Salgado, J. F. (1998). Big five personality dimensions and job performance in Army and civil occupations: A European perspective. *Human Performance*, 11, 271-288.
- SAS Institute Inc. (2001). The SAS system for Windows (Version 8). Cary, NC: SAS Institute Inc.
- Schmidt, F. L. (1988). The problem of group differences in ability scores in employment selection. *Journal of Vocational Behavior*, 33, 272-292.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15, 187-210.
- Schmidt, F. L. & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Schmidt, F. L. & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128-1137.
- Schmidt, F. L. & Hunter, J. E. (1992). Causal modeling of processes determining job performance. Current Directions in Psychological Science, 1, 89-92.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings, *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures of work-force productivity. *Journal of Applied Psychology*, 64, 609-626.
- *Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology*, 73, 46-57.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., Rothstein Hirsh, H., Sackett, P. R., Schmitt, N., Tenopyr, M. L., Kehoe, J., & Zedeck, S. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, 38, 697-798.

- Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology*, 66, 166-185.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter bayseian validity generalization procedure. *Personnel Psychology*, 32, 257-281.
- Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, R., Pearlman, K. & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology*, 78, 3-12.
- *Schmidt, F. L. & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology*, 52, 445-464.
- Schmidt, F. L., Viswesvaran, C. & Ones, D. S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology*, 53, 901-912.
- *Schmitt, N. & Ryan, A. M. R. (1993). The big five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78, 966-974.
- Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82, 719-730.
- Schnake. M. (1991). Organizational citizenship: A review, proposed model, and research agenda. *Human Relations*, 44, 735-759.
- *Schnake. M., Dumler, M. P. & Cochran, D. S. (1993). The relationship between "traditional" leadership, "super" leadership, and organizational citizenship behavior. *Group and Organization Management*, 18, 352-365.
- Scullen, S. E., Mount, M. K. & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, 85, 956-970.
- *Scullen, S. E., Mount, M. K. & Judge, T. A. (2003). Evidence of the construct validity of developmental ratings of managerial performance. *Journal of Applied Psychology*, 88, 50-66.
- Settoon, R. P. & Mossholder, K. W. (2002). Relationship quality and relationship context as antecedents of person- and task-focused interpersonal citizenship behavior. *Journal of Applied Psychology*, 87, 255-267.

- Shaw, J. C., Wild, E. & Colquitt, J. A. (2003). To justify or excuse?: A meta-analytic review of the effects of explanations. *Journal of Applied Psychology*, 88, 444-458.
- *Shore, L. M. & Wayne, S. J. (1993). Commitment and employee behavior: Comparison of affective commitment and continuance commitment with perceived organizational support. *Journal of Applied Psychology*, 78, 774-480.
- *Shore, L. M., Barksdale, K., & Shore, T. H. (1995). Managerial perceptions of employee commitment to the organization. *Academy of Management Journal*, 38, 1593-1615.
- *Shore, L. M., Tetrick, L. E., Shore, T. H., & Barksdale, K. (2000). Construct validity of measures of Becker's side bet theory. *Journal of Vocational Behavior*, 57, 428-444.
- Solomonson, A. L. & Lance, C. E. (1997). Examination of the relationship between true halo and halo error in performance ratings. *Journal of Applied Psychology*, 82, 665-674.
- *Stewart, G. L. (1996). Reward structure as a moderator of the relationship between extraversion and sales performance. *Journal of Applied Psychology*, 81, 619-627.
- *Stokes, G. S., & Searcy, C. A. (1999). Specification of scales in biodata form development: Rational vs. empirical and global vs. specific. *International Journal* of Selection & Assessment, 7, 72-85.
- *Tansky, J. W. (1993). Justice and organizational citizenship behavior: What is the relationship? *Employee Responsibilities and Rights Journal*, 6, 195-207.
- Taylor, F. W. (1911). Scientific management. *The principles of scientific management*, (pp. 30-48; 57-60). New York:Harper & Row.
- Taylor, F. W. (1912). What is scientific management? Excerpted testimony before the U.S. House of Representatives.
- *Tepper, B. J., Lockhart, D. & Hoobler, J. (2001). Justice, citizenship, and role definition effects. *Journal of Applied Psychology*, 86, 789-796.
- Tett, R. P. & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology*, 46, 259-293.
- Thompson, J. D. (1967). Organizations in action. New York: McGraw-Hill.

٩

- *Tompson, H. B., & Werner, J. M. (1997). The impact of role conflict/facilitation on core and discretionary behaviors: Testing a mediated model. *Journal of Management*, 23, 583-601.
- *Turnipseed, D. L. (2002). Are good soldiers good? Exploring the link between organization citizenship behavior and personal ethics. *Journal of Business Research*, 55, 1-15.
- *Turnipseed, D. L. (2003). Hardy personality: A potential link with organizational citizenship behavior. *Psychological Reports*, 93, 529-543.
- *Turnley, W. H., Bolino, M. C., Lester, S. W., & Bloodgood, J. M. (2003). The impact of psychological contract fulfillment on the performance of in-role and organizational citizenship behaviors. *Journal of Management*, 29, 187-206.
- U.S. Department of Labor. (1991). *Dictionary of Occupational Titles* (Rev. 4th ed.). Washington, DC: U.S. Government Printing Office.
- Van Dyne, L., Cummings, L. L., & Parks, J. M. (1995). Extra-role behaviors: In pursuit of construct and definitional clarity (a bridge over muddied waters). *Research in* Organizational Behavior, 17, 215-285.
- *Van Dyne, L., Graham, J. W., & Dienesch, R. M. (1994). Organizational citizenship behavior: Construct redefinition, measurement, and validation. *Academy of Management Journal*, 37, 765-802.
- *Van Scotter, J. R. & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, 81, 525-531.
- Van Scotter, J. R., Motowidlo, S. J., & Cross, T. C. (2000). Effects of task performance and contextual performance on systemic rewards. *Journal of Applied Psychology*, 85, 526-535.
- *VanYperen, N. W., van den Berg, A. E., & Willering, M. C. (1999). Towards a better understanding of the link between participation in decision-making and organizational citizenship behaviour: A multilevel analysis. *Journal of Occupational & Organizational Psychology*, 72, 377-392.
- *Villanova, P., Bernardin, H. J., Johnson, D. L., & Dahmus, S. A. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnomver of motion picture theater personnel. *Personnel Psychology*, 47, 73-90.
- Vinchur, A. J., Schippmann, J. S., Switzer, F. S. III, & Roth, P. L. (1998). A metaanalytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, 83, 586-597.

- Viswesvaran, C. & D. S. Ones (1995). Theory testing: Combining psychometric metaanalysis and structural equations modeling. *Personnel Psychology*, 48, 865-885.
- Viswesvaran, C. & Ones, D. S. (2002). Agreements and disagreements on the role of general mental ability (GMA) in industrial, work, and organizational psychology. *Human Performance*, 15, 212-231.
- Viswesvaran, C., Ones, D. S. & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557-574.
- Wanous, J. P., Sullivan, S. E., & Malinak, J. (1989). The role of judgment calls in metaanalysis. Journal of Applied Psychology, 74, 259-264.
- Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. Journal of Applied Psychology, 75, 315-321.
- *Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17, 601-617.
- *Witt, L. A., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology*, 87, 164-169.
- Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.) *Testing structural equation models*. (pp.256-293) Newbury Park, CA: Sage Publications.
- *Yoon, M. H., & Suh, J. (2003). Organizational citizenship behaviors and service quality as external effectiveness of contact employees. *Journal of Business Research*, 56, 597-611.

