

## ABSTRACT

JOHNSON, EMILY CATHERINE. A Multi-Level Investigation of Overall Job Performance Ratings. (Under the direction of Adam Meade and Mark Wilson.)

Although ratings of overall job performance are a common criterion measure in I/O psychology, research dedicated to understanding how overall performance is conceptualized by the individuals responsible for evaluating performance is limited. In this study, performance data collected from the immediate supervisors of state troopers was used to examine the relationship between dimension ratings and ratings of overall performance. Hierarchical linear modeling analyses indicated that, while both aspects of performance contribute to ratings of overall performance, the relationship between contextual performance and overall ratings varied across supervisors. Hypotheses were offered regarding characteristics of the supervisor and the work context that might explain this variance, but were not supported. Implications for practice and directions for future research are discussed.

A Multi-Level Investigation of Overall Job Performance Ratings.

by  
Emily Catherine Johnson

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Psychology

Raleigh, North Carolina

March 9, 2009

APPROVED BY:

---

Dr. Adam W. Meade  
Committee Chair

---

Dr. Mark A. Wilson  
Committee Co-Chair

---

Dr. S. Bartholomew Craig

---

Dr. Joan J. Michael

## BIOGRAPHY

Emily C. Johnson was born in Lexington, Kentucky to parents Gregory and Marianne Johnson. In 2004, she received a Bachelor of Science degree in Psychology from Tulane University. Although reluctant to leave New Orleans, Emily entered the Industrial/Organizational Psychology program at North Carolina State University in fall of 2004 and went on to earn her Master of Science degree in 2007.

## ACKNOWLEDGMENTS

I'd like to thank the I/O program faculty for providing me with the tools and education that made this dissertation possible. In particular, special thanks to my advisor, Adam Meade, for his invaluable support and guidance during my time at N.C. State and to Mark Wilson for giving me an opportunity to do some very cool work and for getting me excited about performance research in the first place. I can't (and don't want to) imagine what this process would have been like without either of you. I'd also like to acknowledge my advisory committee, Drs. Adam Meade, Mark Wilson, Bart Craig, and Joan Michael, for providing such useful feedback throughout the dissertation process and for being so supportive of my research.

## TABLE OF CONTENTS

LIST OF TABLES.....	vi
A MULTI-LEVEL INVESTIGATION OF OVERALL JOB PERFORMANCE RATINGS .....	1
Introduction .....	1
Task Performance .....	2
Contextual Performance.....	3
Overall Performance .....	4
Level-1 Models .....	6
Other level-1 influences .....	10
Level-2 Models .....	12
Supervisor experience .....	12
Work environment.....	13
Method .....	16
Participants.....	16
Supervisors .....	16
Employees .....	17
Procedure .....	17
Measures .....	18
Task performance .....	18
Contextual performance .....	18
Overall performance.....	19
Performance context.....	19
Other measures .....	19
Analyses.....	20
Results .....	20
Fully Unconditional Model.....	20
Level 1 Models .....	22
Level 2 Models .....	25
Intercepts-as-outcomes model.....	25
Slopes-as-outcomes .....	26
Discussion .....	27
Main Contributions .....	28
Limitations and Future Directions .....	32
Conclusions.....	33
References .....	35
APPENDICES .....	44

Appendix A. List of Trooper Job Tasks.....	45
Appendix B. Contextual Performance Items .....	47
Appendix C. Dissertation Proposal Document Introduction and Method Sections.....	54
Appendix D. Dissertation Proposal Document Tables .....	92

## LIST OF TABLES

Table 1	Fit Indices for Task Performance Models .....	97
Table 2	Results of Task and Contextual Performance Model .....	98

# A MULTI-LEVEL INVESTIGATION OF OVERALL JOB PERFORMANCE RATINGS

## Introduction

Job performance is typically conceptualized as “actions and behaviors that are under the control of the individual that contribute to the goals of the organization,” (Rotundo & Sackett, 2002, p. 66). Especially important to this definition are the ideas that performance is behavior (as opposed to the results of behavior) and that this behavior can be evaluated in terms of its contribution to organizational goals (Campbell, 1990). Campbell (1990) asserts that job performance is inherently multidimensional, a view that is shared by the vast majority of performance researchers (e.g., Borman & Motowidlo, 1993; Campbell, Gasser, & Oswald, 1996; Hesketh & Neal, 1999; Motowidlo, Borman, & Schmit, 1997; Murphy & Shiarella, 1997). However, it is also generally accepted that, while multidimensional conceptions of performance are appropriate when conducting research, for decision making in organizations, a unidimensional, or composite criteria is preferred (Schmidt & Kaplan, 1971). And while the notion that raters will differ in the way they combine information to arrive at an overall rating is far from new (e.g., Naylor & Wherry, 1965), in the job performance domain, research dedicated to explaining this variability is limited.

According to Campbell (1990), the process of assigning a rating to an individual’s job performance is as follows: observation, sampling, encoding, storage, retrieval, evaluation, differential weighting, and composite scoring. In the current study, I confine my focus to the stages of this process in which supervisors combine performance information in order to assign an overall performance rating to their subordinate. Of specific interest is whether



different supervisors use different types of information to inform their ratings of subordinate overall performance and, if so, can these differences be explained by characteristics of the supervisor or work context? In the following sections, I review prevailing theory related to the task and contextual dimensions of performance, as well as overall performance (or, more specifically, *ratings* of overall performance). The remainder of the introduction will describe my specific hypotheses and research questions as well as relevant empirical findings to date.

### *Task Performance*

Task performance refers to behaviors and activities that support the organization's 'technical core,' and can involve the execution of technical processes (transforming raw materials into the goods or services provided by the organization) or the maintenance of those processes, for instance by providing raw materials, distributing products, or through planning and coordination functions (Borman & Motowidlo, 1993; Motowidlo, et al., 1997). For the current study, which was conducted in a law enforcement context, activities that contribute directly to the organization's technical core would include issuing citations and making arrests, while activities such as hiring and training new officers serve to maintain the technical core. To perform job tasks successfully, the employee must have sufficient knowledge of both the task and the context in which it is performed as well as the ability to perform the task when appropriate (Murphy & Shiarella, 1997).

Two distinguishing features of task performance are that the requisite behaviors or activities are usually job-specific and that level of performance tends to be predicted by the performer's level of knowledge, skill, and ability (Borman, 2006). Task activities are

formally acknowledged by the organization as job requirements, often explicitly via job descriptions.

### *Contextual Performance*

Ten years after the introduction of organizational citizenship behavior (OCB; Smith, Organ, & Near, 1983), which is said to have been the result of an interest in the behavioral consequences of job satisfaction (Motowidlo, 2000), the contextual performance construct was introduced. Borman and Motowidlo (1993) believed that performance measures used in selection research and practice were deficient in that they focused solely on task performance, while ignoring activities such as persisting, helping, and endorsing organizational objectives, activities which the authors broadly termed, “contextual performance.” The authors argued that the criterion domain consists of both task performance, behaviors that contribute to the organization’s technical processes, and contextual performance, or behaviors that support the broader psychological and social environment in which that technical core must function (Borman & Motowidlo, 1993). Unlike Smith and colleagues (1983), who confined their OCB construct to those behaviors that are “not absolutely required,” Borman and Motowidlo made no such restriction regarding contextual performance. Rather, for a behavior (required or otherwise) to be considered contextual performance, it must contribute to organizational goals *indirectly*, through the work context.

Despite the difference in definition, the two constructs were found to share many of the same behavioral elements, a fact that quickly became a source of confusion for

researchers (Motowidlo, 2000). Perhaps in response to this concern, Organ (1997) published an article in which he redefined the OCB construct to be synonymous with contextual performance, noting that the earlier requirement that OCBs be extra-role was problematic given that job roles tend to be fuzzy and changeable, and are often negotiated between supervisor and subordinate.

Rather than recommend that OCB be abandoned altogether, though, Organ proposes a solution that amounts to changing the name of the contextual performance to OCB. The problem with the term contextual performance, he argues, is that is it, “cold, gray, and bloodless,” (Organ, 1997, p. 91). However, because the definition of contextual performance has been more consistent over the years and because I believe the term better represents the defining feature of this performance dimension (i.e., that behavior contributes to organizational effectiveness through the work *context*), I adopt this terminology for the remainder of this paper.

### *Overall Performance Ratings*

An individual’s overall performance rating can be thought of as a measure of his or her organizational worth (Motowidlo & Van Scotter, 1994). This is because overall job performance represents the aggregate contribution of an employee’s behaviors to the organization’s goals, over the course of the evaluation period (Motowidlo, et al., 1997). Implicit in this definition is that job performance must include *only* those behaviors that contribute to the organization’s goals (Campbell, 1990). But, which behaviors are these? Campbell explains that there are two judgments involved in this determination. The first is a

*value* judgment as to what the organizational goals actually are. Sometimes organizational goals are formally articulated, but often there is some ambiguity as to what really matters. The second judgment is an *expert* judgment and refers to the decision as to whether or not observed behaviors contribute to the organization's goals.

The model seems to suggest that the rating assigned to an incumbent's overall job performance will depend not only on the level at which he or she performs certain behaviors, but also on the rater's beliefs about the goals of the organization and his or her mental model relating behaviors to goals. If a rater believes that a certain behavior contributes very little to the organizational goals he or she deems valuable, the behavior should have very little influence on the overall rating, even if the incumbent performed exceptionally well in that area.

Related to this is the idea that supervisors hold certain 'role expectations' regarding what their employees are supposed to do. According to role theory, the role expectations that one holds are influenced by characteristics of both the individual and the context in which he or she is embedded (Welbourne, D. E. Johnson, & Erez, 1998). Lievens et al. (2008) suggest that performance evaluation represents an appraisal of how well a worker's behavior conforms to the role expectations held by the rater (typically a supervisor). Thus, if a worker's supervisor does not believe that contextual performance behaviors are an important aspect of that worker's job, those behaviors will not weigh heavily into the supervisor's assessment.

In the current study, I describe a hierarchical investigation into the influence of various factors on ratings of overall job performance. By using a hierarchical linear modeling (HLM) approach, it is possible to investigate factors that are specific to an individual employee, such as his or her time on the job, as well as factors that are shared within groups employees, such as characteristics of the supervisor or of the work environment. In this study, I hypothesize that both individual and situation factors play a role in whether or not supervisors view contextual performance as role-required and, as such, affect its weight in supervisors' evaluations of their employees' performance. In the following sections, I discuss these factors along with my specific hypotheses and research questions. For clarity, I begin with a discussion of relationships at the lowest level (Level 1), or those that involve factors that vary among individual employees, and then turn to higher-level (Level 2) models, where I discuss the possible influence of those factors which vary at the supervisor-level.

#### *Level 1 Models*

The relationship between employees' overall performance ratings and both the task and contextual components of his or her work performance has been well-established through field studies as well as experimental and meta-analytic investigations (Bergeron, 2007; Podsakoff, MacKenzie, Paine, & Bachrach, 2000). Based on the substantial empirical evidence that task performance and contextual performance are distinct constructs (e.g., LePine, Erez, & D. E. Johnson, 2002), both of which contribute to ratings of overall performance (e.g., Borman, White, & Dorsey, 1995; Conway, 1999; Motowidlo & Van Scotter, 1994), I predicted:

*Hypothesis 1:* Ratings of both task performance and contextual performance will be positively related to overall performance ratings.

In addition, because Bergeron's (2007) review of published studies indicates that task performance is usually, but not always, given more weight in overall performance ratings, I also compared the variance in overall performance ratings that can be attributed to task and contextual performance ratings:

*Research Question 1:* What proportion of the variance in overall job performance ratings can be attributed to task vs. contextual performance?

One limitation shared by many previous studies is that they ignore the possibility that raters have different ways of using information about task and contextual performance when assigning overall performance ratings. That is, while previous studies have demonstrated that, across raters, ratings of both task and contextual performance are related to overall performance ratings, few attempts have been made to determine whether these relationships vary, perhaps as a function of the rater or situation.

In efforts to overcome this limitation, some more recent studies have adopted an experimental policy-capturing methodology. In these studies, performance profiles are created in which the performance of a hypothetical employee is described (i.e., a paper person), where levels of task and contextual performance are manipulated. Using the performance ratings assigned to the paper people by study participants, a "captured rating policy" equation can be generated for each rater. This equation describes the way the individual rater weights performance dimensions when assessing overall performance (J. W.

Johnson, 2001; Rotundo & Sackett, 2002). Because policy-capturing studies result in a separate equation for each rater, they overcome some of the limitations associated with other performance appraisal studies, by providing more in-depth information about the differences between individuals (Karren & Barringer, 2002). However, the experimental nature of policy-capturing studies presents a different set of limitations. Rating the described behavior of a paper person clearly involves a fundamentally different cognitive process compared to the task of rating actual, observed performance of real-world employees, in part because performance profiles typically contain very little irrelevant or ambiguous information (Landy, 2008; Murphy, Herr, Lockhart, & Maguire, 1986). Laboratory studies also fail to account for the on-going relationship between supervisor and employee or other potentially influential aspects of the political, social, and affective context in which performance appraisal occurs (Bretz, Milkovich, & Read, 1992).

Limitations aside, the results of these studies suggest that raters likely vary in the way they combine information about performance dimensions into ratings of overall performance. For example, using the policy capturing approach, Rotundo and Sackett (2002) demonstrated that raters do not share a common policy for weighting performance dimensions when they evaluate overall performance. Building on this research, Lievens and colleagues (2008) found that their organizational culture manipulation (high vs. low team-based culture) accounted for significant observed variance in raters' policies for combining performance dimensions. Research is needed to determine whether or not these findings do generalize to the field and, if so, to investigate what other factors may have an influence.

The current study extends previous research by employing HLM to investigate whether various characteristics of the subordinate (ratee), supervisor (rater), and work environment influence the relationships between performance dimensions and overall performance ratings. As in policy-capturing studies, each supervisor rated multiple employees. Also similar to the policy-capturing approach, the HLM approach involves the estimation of a separate equation for each rater/supervisor. However, unlike the policy capturing approach, the current study does not rely on experimental methodology, but rather involved real employees in a real work environment.

A number of authors (e.g., Bergeron, 2007; Murphy & Cleveland, 1995; Organ, 1997) have suggested that, while not always the case, it is often useful to think of task performance as those behaviors and activities listed in an employee's formal job description. When characteristics of effective task performance are formally prescribed by the organization, as is the case with formal job descriptions, one might expect that the importance of such behaviors will be understood by all (or most of) the relevant organizational members. In this case, these activities and behaviors are publically recognized by the organization as defining aspects of the job, or what the employee has been hired to do and will be evaluated for doing. Accordingly, I argue that supervisors will view the contribution of task behaviors to overall performance in a similar manner.

*Hypothesis 2(a):* The strength of the relationship between task performance and overall performance will not vary across supervisors.



Compared to task performance, the organizational policy regarding contextual performance is likely less clear. Even though the prevailing definition does not limit contextual performance to discretionary behaviors, contextual behaviors are less likely than task behaviors to be formally required or clearly linked to rewards (Organ, 1997). As the contribution of contextual performance behaviors to overall performance is not generally expected to be formalized by the organization, the possibility exists that, within an organization, individuals will vary in their beliefs about contextual performance. In support of this possibility, Morrison (1994) found that employees' beliefs about whether or not various contextual performance items were part of their job role varied within a single organization. If employees differ in the extent to which they see contextual behaviors as part of the job they are hired to do, supervisors might also. Unlike the task performance-overall performance relationship, which I argue is communicated by the organization and therefore will not vary across supervisors, the link between contextual performance and overall performance may be largely determined by the implicit theory of the individual. If this is the case, the slope relating subordinates' contextual performance to overall performance would vary across the supervisors providing those ratings. In other words:

*Hypothesis 2(b):* The strength of the relationship between contextual performance and overall performance ratings will vary across supervisors.

*Other level-1 influences.* Another subordinate-level variable that might influence ratings of overall performance is work experience. Work experience has been defined as the events an employee perceives or experiences that relate to the performance of his or her job

(Quinones, Ford, & Teachout, 1995). In practice, work experience has been operationalized as time on the job (e.g., Schmidt, Hunter, & Outbridge, 1986), time in the organization (e.g., Potter & Fiedler, 1981), time in the occupation (e.g., McDaniel, Schmidt, & Hunter, 1988), and number of times performing a given task (e.g., Spiker, Harper, & Hayes, 1985), to name a few. Quinones and colleagues' (1995) meta-analysis demonstrated that, regardless of the operationalization used, work experience tends to be positively related to overall job performance (estimated  $\rho = .27$ ).

*Hypothesis 3:* Employee (trooper) time on the job will be positively related to overall performance ratings.

Moreover, J. W. Johnson (2001) suggests that an employee's level of experience may influence how his or her supervisor weights the dimensions. I suspect that, at least in some organizations, employee work experience moderates the influence of contextual performance on assessments of overall performance. Organ (1988) for instance, theorized that as individuals move up the organizational ranks, the breadth of their roles expands. I predict that role expansion will similarly occur as employee tenure increases, such that an employee's contextual performance will be weighted more heavily as that employee becomes more experienced. As the proposed sample is drawn from an organization where internal promotion is the norm, high tenure employees may be expected to take on expanded job roles in anticipation of a promotion to a higher rank.

*Hypothesis 4:* Employee (trooper) time on the job will moderate the relationship between contextual performance and overall performance ratings, such that the relationship is stronger for those with more (vs. less) time on the job.

Judge and Ferris (1993) argue that, due to prevalence of factors motivating supervisors not to assign low performance ratings (e.g., lawsuits, paperwork associated with complaints, resentment), the performance rating process is akin to a search for positive information. In this case, opportunity to observe employees' performance would be positively related to overall performance ratings. However, in high stakes jobs such as law enforcement, a supervisor must weigh the advantages of inflating performance ratings against the potential risks of doing so. Failure to document performance deficiencies could result in public or officer safety issues, which might also incur complaints or lawsuits.

*Research Question 2:* Is supervisors' opportunity to observe employee job performance related to overall performance ratings?

### *Level 2 Models*

*Supervisor experience.* Research findings regarding supervisor experience have been mixed, with some studies reporting that supervisors with more tenure tend to be more lenient, and others finding no effect (Landy & Farr, 1980). Judge and Ferris (1993) speculated that with increased tenure comes decreased need to prove one's toughness via harsh ratings and a better grasp on the potential costs of assigning unfavorable ratings, but ultimately failed to find evidence of a link between supervisor experience and overall performance ratings.

*Research Question 3:* Will supervisor experience (time in current rank) positively influence overall performance ratings?

Because the current study was conducted in an organization where internal promotion is the norm, less experienced supervisors were likely asked to rate the performance of employees with whom, until recently, they worked alongside as a peer. Research comparing peers and supervisors (e.g., Conway, 1999; Lievens, et al., 2008; Vey & Campbell, 1994) suggests that the relationship between contextual performance and overall performance ratings might be stronger when ratings are provided by less experienced supervisors, to the extent that these supervisors still have somewhat peer-like associations with the employees they rate. An alternative possibility is that relatively inexperienced supervisors may experience insecurity about their new role and, as a result, focus their performance management efforts around those behaviors and activities that are formally prescribed by the organization (i.e., task performance), to the exclusion of contextual performance (Befort & Hatrup, 2003). Because compelling arguments could be made for both positive and negative cross-level moderation effects, supervisor experience was investigated as a research question.

*Research Question 4:* Does the relationship between contextual and overall performance ratings depend on how long the supervisor providing those ratings has held the rank of sergeant?

*Work environment.* As discussed previously, when performance is evaluated as the contribution of an individual's behaviors to the goals of the organization, the value of those goals must be judged by the organization or, more realistically, certain organizational

members (Campbell, 1990). While Campbell does not speculate as to what types of information might form the basis of this value judgment, characteristics of the environment in which the organization operates would seem relevant. Murphy and Cleveland (1995) hypothesize that characteristics of the environment likely affect performance appraisal through a number of mediating variables, including standards for performance and performance dimensions. Regarding performance dimensions, they write, “the definition of what constitutes performance and the relative importance of different parts of the performance domain are likely to be affected by the environment,” (p. 39). Unfortunately, as the authors go on to note, research addressing the effects of the environment on performance appraisal is “virtually nonexistent,” (p. 43). More recently, Hanson and Borman (2006) lament that “little is known about the mechanisms by which citizenship performance affects organizational effectiveness or supervisor ratings,” but they tentatively suggest that job or situational characteristics may moderate the relationship (p. 166).

Returning to the distinction between the task and contextual dimensions of job performance, researchers generally believe that task activities are performed because they are explicitly required by the job, while the performance of contextual activities tends to have a more to do with an employee’s motivation. However, the assumption that any individual who is motivated to engage in contextual performance behaviors will do so may not always be accurate (Griffin, Neal, & Neale, 2000). Rather, aspects of the work situation, such as task demands, may affect opportunities to perform the more discretionary aspects of the job. Griffin and colleagues suggest that, when task demands are high, the majority of an

individual's time and energy will be consumed by performing task activities, leaving few resources available for contextual behaviors. As Bergeron (2007) puts it, "Recognizing that individuals have limited time, any time spent on OCB comes at the expense of task performance," (p. 1078).

Along these lines, Smith and colleagues' (1983) seminal study found that that individuals with rural (vs. urban) backgrounds are more likely to engage in altruistic behaviors. The authors explain that, "people from urban settings learn to cope with overstimulation from the physical and social environment by screening out many social stimuli, thus decreasing their sensitivity to others' needs," (p. 662). The current study proposes to incorporate this notion of urban versus rural work setting by including two variables expected to shape the work environment, namely, population and number of interstate highways. An environment characterized by more highways and a higher population should mean increased stimulation from the physical and social environments, and, accordingly, increased task demands. In the Griffin study, the authors found that for a sample of air traffic controllers, supervisors' ratings of subordinate task performance were related to ratings of overall effectiveness across situations, while contextual performance (teamwork, professionalism, and supporting organizational objectives) was only related to effectiveness in situations in which task demands were low (based on variables such as traffic volume and traffic complexity). Moreover, if working in an urban district causes supervisors to screen out interpersonal stimuli, as Smith and colleagues suggest, urban supervisors would

probably assign less importance to the social context of work and, therefore, struggle to see the value of contextual performance.

*Hypothesis 5:* The relationship between contextual performance and overall performance ratings will be stronger (vs. weaker) for districts that contain a lower (vs. higher) number of interstates.

*Hypothesis 6:* The relationship between contextual performance and overall performance ratings will be stronger (vs. weaker) for districts with lower (vs. higher) population.

In the following section, I discuss the sample, measures, and analytical procedure used to test these hypotheses and research questions.

## Method

### *Participants*

The sample consisted of North Carolina State Highway Patrol (NCSHP) troopers and their supervising sergeants selected to participate in a recent selection system revalidation effort (see E. C. Johnson, DuVernet, & Wilson, 2008).

*Supervisors.* Only those supervising sergeants who provided usable performance ratings for at least two troopers were included in the analyses. Among those sergeants ( $N = 86$ ), the number of troopers rated ranged from 2 to 13, with a median = 3 ( $M = 3.72$ ,  $SD = 2.07$ ). The majority reported that, at the time of rating, they had been in the role of sergeant for over three years. While no demographic information is available for the supervisor sample, there is no reason to expect that they differed substantially from the total sergeant

population ( $N = 222$ ), which was 16.2% African American, 1.8% American Indian, 81.5% Caucasian, and 96.4% male at the time of data collection.

*Employees.* Usable performance ratings were available for  $N = 324$  troopers. Time between being sworn in as a trooper and the date of the evaluation ranged from 1.96 years to 7.81 years ( $M = 4.85$ ,  $SD = 1.85$ ). The sample was approximately 98% male, with Caucasians making up 86% of the total and African Americans comprising an additional 9%.

### *Procedure*

All performance measures were collected online as part of a selection system validation effort (see E. C. Johnson et al., 2008). Sergeants responsible for supervising the troopers in the sample received a memo which directed them to the rating website and provided them with a list of the names and identification codes for the troopers they were assigned to rate. Sergeants who had supervised a trooper for fewer than six months were instructed to contact the trooper's previous supervisor and become familiar with the trooper's previous performance documentation before completing the ratings.

The online rating website instructed raters that all ratings were being collected for research purposes only and that no ratings would be seen by any member of the NCSHP. Raters entered a "trooper identification code" to indicate the particular trooper they were rating, and then made ratings on 46 items pertaining to the trooper's current levels of task performance, contextual performance, and overall performance.



## *Measures*

*Task performance.* Supervisors rated troopers' performance on 13 job tasks derived from a recent comprehensive job analysis (see Wilson, E. C. Johnson, & DuVernet, 2008), using a 7-point scale which ranged from "poor" to "excellent." A list of the tasks can be found in Appendix A. Task Performance scores were computed as the average rating across the 13 tasks ( $\alpha=.93$ ).

Based on recommendations that predictor variables be centered before being entered into hierarchical analyses, each trooper's raw task performance score were centered using the grand-mean centering approach, wherein the mean score of the entire sample is subtracted from the individual's score. Centering eases interpretation of model coefficients (i.e., in a model in which task performance predicts overall performance, the intercept would be interpreted as the overall performance rating for a trooper with average task performance) and may also improve model estimation (Kreft, de Leeuw, & Aiken, 1995).

*Contextual performance.* Contextual performance ratings were collected using the same online rating form and were calculated as the average rating across 19 items ( $\alpha=.94$ ) designed to assess four of dimensions of citizenship performance proposed by Organ (1988): Altruism, Conscientiousness, Civic Virtue, and Courtesy (Podsakoff, et al., 1990).

Supervisors indicated the degree to which they believed statements were true of the trooper they were evaluating, using a 7-point scale, which ranged from "strongly agree" to "strongly disagree." The scale items can be found in Appendix C. Contextual performance scores were centered using the same grand-mean centering technique described for task performance.

*Overall performance.* After responding to the task performance items but before the contextual performance items, supervisors also provided ratings of troopers' overall performance, by responding to a single item, "After considering everything you know about this Trooper, how would you rate his/her overall performance?" Ratings were made using a 7-point scale, where 1 indicates "poor" and 7 indicates "excellent." The actual ratings provided ranged from 2 to 7 ( $M = 4.99$ ,  $SD = .71$ ).

*Performance context.* The NCSHP is organized geographically into 8 troops, which are further divided into districts. While higher ranking members of the patrol may be responsible for multiple districts, troopers and sergeants are typically assigned to work in one specific district, which can be made up of 1 to 5 counties. In the current study, usable performance ratings were available from supervisors in 62 different districts, and work context was described as the population and number of highways for the particular district in which a supervisor (and his or her employees) work.

*Other measures.* In addition to making the above described performance assessments, the supervising sergeants also provided information about how long they had supervised the trooper they were rating (*"How long have you supervised this Trooper?"*) and how often they have the opportunity to observe the trooper's performance (*"How often do you get to directly observe this Trooper doing their job?"*). Approximately half of the supervisors reported that they had supervised the trooper for at least 2 years and two thirds reported observing the trooper's performance weekly.

## *Analyses*

The variables described above can be summarized as characteristics of troopers, characteristics of the sergeants who supervise them, and characteristics of the work environment. Importantly, these variables exist at multiple levels of analysis: there are variables describing the troopers, and those troopers are grouped within the sergeants who supervise them and evaluate their performance. For data such as this, the most appropriate analytical tool is hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002). The HLM approach allows group- and individual-level residuals to be modeled separately, recognizing that ratings of different employees made by the same supervisor may not be completely independent. Additionally, the HLM methodology allows hypotheses regarding cross-level effects to be tested (e.g., how supervisor characteristics affect the trooper-level relationship between contextual and overall performance ratings). The following section describes the models tested and the results of those tests.

## **Results**

### *Fully Unconditional Model*

The simplest possible hierarchical model, the fully unconditional or “null” model, is typically assessed as the first step in HLM analysis (Raudenbush & Bryk, 2002). The model is said to be fully unconditional in the sense that no predictor variables are entered into the equations at either level. For the present study, where ratings of individual troopers are grouped by the supervising sergeant providing the rating, the unconditional model is written as:

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + r_{ij} \quad (1)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}. \quad (2)$$

In this model, the dependent variable is the rating of individual troopers' overall performance. At level 1, this rating is expressed as the sum of an intercept for supervisor  $j$  ( $\beta_{0j}$ ) and a random effect ( $r_{ij}$ ) associated with the  $i^{\text{th}}$  trooper rated by the  $j^{\text{th}}$  supervisor. At level 2, the intercept is expressed as the sum of the grand-mean rating ( $\gamma_{00}$ ) and a random effect ( $u_{0j}$ ) associated with the  $j^{\text{th}}$  supervisor.

The results of the unconditional model indicated that ratings of overall performance are comprised of both a significant between-supervisor variance component ( $\tau_{00} = 0.10$ ,  $z = 3.01$ ,  $p < .01$ ), as well as a significant within-supervisor component ( $\sigma^2 = 0.40$ ,  $z = 11.04$ ,  $p < .01$ ). That the variance is statistically significant at both levels indicates that the development of models that include individual- and group-level predictors of overall performance is justified. Before turning to those models, however, it is common practice to use these variance components to estimate the intraclass correlation coefficient, calculated as  $\rho = \tau_{00} / (\tau_{00} + \sigma^2)$ . The ICC provides an index of the between-supervisor variance (i.e., variance between level-2 units) as a proportion of the total variability. In the current study, dividing the between-supervisor variance by the total variance yields an ICC equal to .203, which is to say that 20.3% of variability in ratings of overall performance occurs at the supervisor level, while 79.7% (i.e.,  $1 - \rho$ ) occurs at the individual trooper level.

#### *Level-1 Models*

While Hypothesis 1 specifies main effects for two level-1 predictors, task performance and contextual performance, the model was tested using the recommended “step-up” approach (Raudenbush & Bryk, 2002), in which the analyst first tests a single-predictor model and then proceeds by adding variables one at a time. For all of the models described below, if a variable did not have a significant fixed effect and there was no evidence of slope heterogeneity, it was omitted from further models.

To test whether ratings of task performance have a significant, positive relationship with overall performance ratings, as specified by Hypothesis 1, the following model was tested:

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + r_{ij} \quad (3)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (4)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}. \quad (5)$$

The results of this model revealed that the estimated average slope relating trooper task performance and overall performance,  $\gamma_{10}$ , is significantly different from 0 in the hypothesized direction ( $\gamma_{10} = .96$ ,  $t(237) = 25.64$ ,  $p < .01$ ), thereby providing partial support for Hypothesis 1.

To test the hypothesis that contextual performance (CP) is also related to overall performance, a model was tested in which contextual performance replaced task performance as the only level-1 predictor:

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + \beta_{1j}(CP) + r_{ij} \quad (6)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (7)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}. \quad (8)$$

The results of this model revealed that the estimated average slope relating trooper contextual performance and overall performance,  $\gamma_{10}$ , is also significant in the hypothesized direction ( $\gamma_{10} = .55$ ,  $t(237) = 9.87$ ,  $p < .01$ ). Taken together with the results of the task performance model, this result provides full support for Hypothesis 1.

Research Question 1, which concerns the variance in overall performance ratings attributable to task and contextual performance was tested using the following calculation:

$$R^2_{\text{within}} = \frac{\sigma_{uc}^2 - \sigma_c^2}{\sigma_{uc}^2}, \quad (9)$$

where  $\sigma_{uc}^2$  is the residual level-1 variance in the unconditional model (i.e., the model with no predictors) and  $\sigma_c^2$  is the residual level-1 variance for the model which includes the predictor of interest. The results indicate that the percentage of level-1 variability in overall performance that can be attributed to task performance ratings alone is 64.69% while the variability accounted for by contextual performance alone is 54.02%.

As predicted by Hypothesis 2(a), the results of the task performance model indicated that the slope relating task performance to overall performance does not vary significantly across supervisors ( $\tau_{11} = .004$ ,  $Z = .42$ ,  $p = .34$ ). In addition, inspection of the various fit indices associated with the two models (see Table 1) provides additional support for constraining the slope across supervisors. Conversely, as predicted by Hypothesis 2(b), the results of the contextual performance model indicated that the slope relating contextual

performance to overall performance does vary significantly across supervisors ( $\tau_{11} = .11$ ,  $Z = 2.91$ ,  $p < .01$ ). Based on these results, a model was specified in which both task performance and contextual performance are included as level-1 predictors of overall performance, with the contextual performance slope, but not the task performance slope, allowed to vary across level-2 units. The results of this model indicate that, together, task and contextual performance account for 73.52% of the trooper-level variability in overall performance ratings. The parameter estimates for this model can be found in Table 2. Note that the level-1 residual variance remained significant, even after the addition of task and contextual performance ( $\sigma^2 = 0.11$ ,  $z = 9.79$ ,  $p < .01$ ). Thus, additional potential level-1 predictors of overall performance ratings were investigated, as described below.

Hypotheses 3 and 4 stated that trooper tenure would be positively related to overall performance and that tenure will moderate the relationship between contextual performance and overall performance. The model to test these hypotheses is:

$$\text{Level 1:} \quad (10)$$

$$Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + \beta_{3j}(tenure) + \beta_{4j}(CP * tenure) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (11)$$

$$\beta_{1j} = \gamma_{10} \quad (12)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (13)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \quad (14)$$

$$\beta_{4j} = \gamma_{40} + u_{4j} \quad (15)$$

The results of this model indicate that trooper tenure does not have a significant main effect on performance ratings ( $\gamma_{30} = .00$ ,  $t(234) = .72$ ,  $p = .47$ ) and also does not significantly moderate the relationship between contextual performance and overall performance ( $\gamma_{40} = .00$ ,  $t(234) = .6$ ,  $p = .55$ ). Thus, no support was found for Hypothesis 3 or 4.

To answer Research Question 2, which asks whether a supervisor's opportunity to observe a trooper will be related to overall performance ratings, a model was tested in which trooper tenure was removed from the above model and opportunity to observe was added. The results of this model indicate that opportunity to observe has no appreciable effect on overall performance ratings ( $\gamma_{30} = .01$ ,  $t(235) = .34$ ,  $p = .74$ ).

#### *Level-2 Models*

*Intercepts-as-outcomes model.* Research Question 3 specified a cross-level main effect, namely, that a supervisor's amount of experience (operationalized as self-reported time in current rank) would be positively related to ratings of overall performance. Whereas the models tested up to this point have been unconditional at level 2 (i.e., no attempt was made to explain variance in slopes or intercepts), here the level-2 predictor, supervisor experience, is added to the intercept equation (Equation 17):

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + r_{ij} \quad (16)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{experience}) + u_{0j} \quad (17)$$

$$\beta_{1j} = \gamma_{10} \quad (18)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (19)$$



The results of this model indicated that supervisor experience does not significantly affect overall performance ratings ( $\gamma_{01} = -.01$ ,  $t(84) = -.47$ ,  $p = .64$ ).

*Slopes-as-outcomes model.* For Research Question 4, which asked whether the relationship between contextual performance and overall performance depends on supervisor experience, the intercepts-as-outcomes model was modified to include experience in the contextual performance slope equation (Equation 23) as well as the intercept equation (Equation 21):

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + r_{ij} \quad (20)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{experience}) + u_{0j} \quad (21)$$

$$\beta_{1j} = \gamma_{10} \quad (22)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{experience}) + u_{2j} \quad (23)$$

The results of this model indicated that supervisor experience does not moderate the relationship between ratings of contextual and ratings of overall performance ( $\gamma_{21} = -.01$ ,  $t(235) = -.36$ ,  $p = .72$ ).

The final hypotheses asserted that characteristics of the district a particular supervisor is responsible for (namely, the district's number of interstates and total population) would moderate the contextual-overall performance relationship. To test Hypothesis 5, regarding interstates, the following model was tested:

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + r_{ij} \quad (24)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{interstates}) + u_{0j} \quad (25)$$

$$\beta_{1j} = \gamma_{10} \quad (26)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{interstates}) + u_{2j}. \quad (27)$$

The results of this model indicated that number of highways does not moderate the relationship between ratings of contextual and ratings of overall performance ( $\gamma_{21} = -.05$ ,  $t(235) = -.99$ ,  $p = .32$ ), nor does it have a significant main effect on overall performance ratings ( $\gamma_{01} = -.02$ ,  $t(84) = -.45$ ,  $p = .66$ ). To test Hypothesis 6, regarding district population, a model nearly identical to the interstate model was tested, where the non-significant interstates variable was replaced with the population variable. The results of this model indicated that, as with interstates, population does not moderate the relationship between ratings of contextual and ratings of overall performance ( $\gamma_{21} = -.02$ ,  $t(235) = -.48$ ,  $p = .63$ ), nor did it have a significant main effect on overall performance ratings ( $\gamma_{01} = -.05$ ,  $t(84) = -1.72$ ,  $p = .09$ ).

## Discussion

The goal of this study was to shed light on how factors at multiple levels of analysis might influence assessments of overall performance. To this end, a series of HLM models was tested, where state troopers were nested within their supervising sergeants. Task and contextual performance ratings were both found to be positively related to overall performance ratings, with the strength of the relationship between contextual performance and overall performance varying across supervisors.

Hypotheses 1 and 2 stated that both task and contextual dimensions of performance would be significantly related to overall ratings, but that only the contextual performance

slope would vary at the supervisor level. These hypotheses were supported. Moreover, though both aspects of performance were found to inform ratings of overall performance, in answer to Research Question 1, task performance accounted for the greater proportion of variance.

Hypotheses 3 and 4 proposed that employee work experience would be positively related to ratings of overall performance and would moderate the relationship between contextual performance and overall ratings. Neither of these hypotheses was supported. Similarly, no evidence of a relationship between supervisor experience and overall ratings (Research Question 3) was found. Supervisor opportunity to observe employee performance also did not seem to influence overall performance ratings (Research Question 2).

Three level-2 variables were investigated as moderators of the contextual performance-overall performance slope: supervisor experience, number of interstates in the district, and district population (Research Question 4, Hypotheses 5 and 6, respectively). However, none of the hypothesized cross-level moderation relationships were supported.

### *Main Contributions*

This study makes a number of contributions to the job performance literature. The results indicated that, within a single organization, the relationship between task performance and overall performance ratings is generally consistent across raters. All subordinates rated in the current study shared a common job description and thus were expected to perform, and rewarded for performing, identical job duties. Accordingly, I reasoned that supervisors would share a common set of expectations regarding the task requirements of the job and, as such,

evaluate task performance as relating to overall performance in largely the same way. The finding that the task performance-overall rating slope did not vary significantly across raters is supportive of this line of reasoning.

On the other hand, the relationship between contextual performance and overall performance varied significantly across raters. Contextual performance behaviors are far less likely than task behaviors to be formally acknowledged by the organization as job requirements (Organ, 1997). Consequently, supervisors receive less information from the organization regarding how and to what extent their employees' contextual performance is related to successful job performance. The finding of variable slopes relating contextual performance to overall performance ratings at the supervisor level suggests that, unlike task performance, beliefs about the contribution of contextual performance are idiosyncratic to individual supervisors. Thus, it may be the case that, in the absence of information from the organization, evaluations of employee job performance are informed by the rater's own implicit theories regarding contextual performance. Taken together, these findings suggest an interesting distinction between the two dimensions of performance. Namely, perhaps the value of task behaviors, in terms of their contribution to the organization's overall goals, is understood across members of the organization, while the perceived value of contextual behavior is more idiosyncratic to individual organizational members.

The current study attempted to identify factors that might drive these rater-level differences regarding the contextual performance-overall performance relationship by examining supervisors' time in current job role as well as characteristics of the work

environment, namely, population and number of highways. None of these variables explained significant variability in the supervisor slopes. This result is troubling as, without knowing what factors influence how raters use contextual performance information in their overall assessment, researchers and practitioners alike have no way of knowing when performance evaluations are based on a common understanding of overall job performance and when they are not.

For researchers, the notion that different individuals may use performance information differently when rating overall performance has some serious implications for the use of overall performance ratings as criteria in test validation research. Murphy and Shiarella (1997) demonstrated that the validity of selection tests can vary considerably depending on how performance dimensions are weighted to form an overall performance composite. In addition, procedures such as meta-analysis assume that measures of overall performance assess the same construct across studies. In the absence of research that identifies specific characteristics of raters and work context that might be expected to affect overall evaluations, researchers should proceed with caution when attempting to draw conclusions from research which relies on overall performance ratings as a criterion measure.

Within organizations, interventions might be designed with the goal of developing the same type of shared understanding as to how contextual behaviors relate to employees' performance of their job roles as is currently apparent regarding task behaviors. This would require increased formalization of the contextual performance requirements of jobs, through job descriptions and appraisal tools. Providing rater training on the new instruments would

increase the salience of such behaviors and enhance the likelihood that supervisors evaluate their employees using the same standards for successful performance. Alternatively, the use of raters might be circumvented entirely in determinations of overall performance through the use of statistical procedures for combining dimension ratings.

An additional aim of the current study was to highlight the potential benefits of the HLM approach for job performance research. This methodology allows multiple levels of influences on job performance ratings to be investigated, thus providing a more complete picture of the performance appraisal process. HLM can serve as an invaluable tool for furthering our understanding of performance appraisal as a process that is influenced not only by the employee being evaluated, but also by the employee's supervisor and the performance context. Moreover, hierarchical models allow that ratings of different employees made by the same supervisor may be more similar than ratings of different employees made by different supervisors, by estimating separate residuals at the employee and supervisor level of analysis. When collecting performance data in organizations, departmental structure often simply does not allow each supervisor to rate only a single employee (Judge & Ferris, 1993). When employees are nested within supervisors or work contexts, HLM provides a more statistically sound approach to analyzing performance data, as compared to OLS regression, which assumes independence of observations (Raudenbush & Bryk, 2002).

#### *Limitations and Future Directions*

Future research should address rater and contextual characteristics that influence the contextual performance-overall performance relationship. The finding of significant level-2

variance in slopes indicates that this is a promising direction for research. Although the current study attempted to explain this variance, none of the hypotheses were supported. One reason for this failure to find a relationship may be that the variables included were simply not the correct variables. For example, number of interstates and district population may not adequately represent the task demands of the work environment. Likewise, supervisor time in rank fails to capture specific experiences that may have important influences on how supervisors come to think about the role of contextual performance. To this end, measuring work environment demands more directly or variables such as supervisor personality or leadership style may provide interesting results.

Another reason for the observed results may be the use of a single item rating of overall performance. Because the single-item measures of job performance are widely used in decision making in organizations (Judge & Ferris, 1993; Wanous & Hudy, 2001), the single item approach was adopted in order to most accurately reflect the cognitive task of rating overall performance as it would be encountered in practice, thus bolstering the external validity of the study. However, in the research literature, single-item measures tend to be criticized on the grounds that their reliability cannot be estimated and that, if it could be, it would be unacceptably low (Wanous & Hudy, 2001). In one attempt to determine the test-retest reliability of a single-item measure of performance, Judge and Ferris (1993) analyzed student ratings of paper people, yielding a test-retest coefficient of .67. However, it is hard to say how their finding relates to ratings of observed performance by a non-student sample. To the extent that performance ratings in the present study contain measurement error,

relationships between overall performance and other study variables will be attenuated. Future studies may consider utilizing multi-item measures of overall performance, thus increasing the likelihood that the construct is measured reliably.

Finally, it remains to be seen whether the conclusions drawn from the present study are specific to the law enforcement population or whether they can be supported across work environments. Because the notion that contextual performance tends to be less clearly linked to organizational rewards than is task performance (Organ, 1997) is not a law enforcement-specific, it seems likely that the slope relating contextual performance to overall performance varies across raters in a variety of organizational settings. Similarly, to the extent to which organizations send clear messages about the task demands of a job, one would expect to find the same sort of consistency in task-overall performance slopes observed in the present sample. However, further research is needed to investigate the generalizability of the current study's findings.

### *Conclusions*

Recent studies suggest that raters are guided by unique policies regarding how different aspects of performance should be combined to produce an overall rating (e.g., Lievens et al., 2008; Rotundo & Sackett, 2002). The current study builds on existing research by using a multi-level approach to show that the relationship between contextual performance and overall performance is more profoundly affected by the individual rater differences than is the relationship between task performance and overall performance. Additionally, the HLM approach used in the current study overcame many limitations of



policy capturing studies, which involve rating the performance of hypothetical employees (see Landy, 2008; Murphy, et al., 1986), by allowing the effects of real organizational forces to be investigated. Research is needed to examine how individual and contextual factors influence the weight given to contextual performance as well as to determine whether there are organizational factors (such as socialization or culture) that increase the likelihood that raters will perceive the contribution of task behaviors to overall job performance similarly.

## REFERENCES

- Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence* , 10, 109-139.
- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin* , 102, 3-27.
- 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.
- Befort, N., & Hattrup, K. (2003). Valuing task and contextual performance: Experience, job roles, and ratings of the importance of job behaviors. *Applied H. R. M. Research* , 8, 17-32.
- Bergeron, D. M. (2007). The potential paradox of organizational citizenship behavior: Good citizens at what cost? *Academy of Management Review* , 32, 1078-1096.
- Borman, W. C. (2006). The Concept of Organizational Citizenship Behavior. *Current Directions in Psychological Science* , 13, 238-241.
- 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., White, L. A., & Dorsey, D. W. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology* , 80, 168-177.
- Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management* , 18, 321-352.
- 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., 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: Jossey-Bass Publishers.
- Coleman, V. I., & Borman, W. C. (2000). Investigating the underlying structure of the citizenship performance domain. *Human Resource Management Review* , 10, 25-44.
- Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology* , 84, 3-13.
- Cortina, J. M. (2002). Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management* , 28, 339-362.

- Griffin, M. A., Neal, A., & Neale, M. (2000). The contribution of task performance and contextual performance to effectiveness: Investigating the role of situational constraints. *Applied Psychology: An International Review* , 49, 517-533.
- Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal* , 50 (2), 327-347.
- Hanson, M. A., & Borman, W. C. (2006). Citizenship performance: An integrative review and motivational analysis. In W. Bennet Jr., C. E. Lance, & D. J. Woehr (Eds.), *Performance measurement: Current perspectives and future challenges* (pp. 141-173). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hesketh, B., & Neal, A. (1999). Technology and performance. In D. R. Ilgen, & E. D. Pulakos (Eds.), *The changing nature of performance: Implications for staffing, motivation, and development* (pp. 21-55). San Francisco: Jossey-Bass.
- Hoffman, B. J., Blair, C. A., Meriac, J. P., & Woehr, D. J. (2007). Expanding the criterion domain? A quantitative review of the OCB literature. *Journal of Applied Psychology* , 92, 555-566.
- Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear modeling. *Journal of Management* , 23 (6), 623-644.
- Ilies, R., Nahrgang, J. D., & Morgeson, F. P. (2007). Leader-member exchange and citizenship behaviors: A meta-analysis. *Journal of Applied Psychology* , 92, 269-277.

- Johnson, E. C., DuVernet, A. M., & Wilson, M. A. (2008). *Validation of a Trooper Selection System: Project Technical Report*. Raleigh, NC: Mark Wilson Associates.
- 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., & Ferris, G. R. (1993). Social context of performance evaluation decisions. *Academy of Management Journal* , 36, 80-105.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement* , 20, 141-151.
- Kamdar, D., McAllister, D. J., & Turban, D. B. (2006). 'All in a day's work': How follower individual differences and justice perceptions predict OCB role definition and behavior. *Journal of Applied Psychology* , 91, 841-855.
- Karren, R. J., & Barringer, M. W. (2002). A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods* , 5, 337-361.
- Klein, K. J., & Kozlowski, S. W. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods* , 3 (3), 211-236.
- Kreft, I. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in heirarchical linear models. *Multivariate Behavioral Research* , 30, 1-21.

- Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology* , 1 (4), 379-392.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin* , 87, 72-107.
- 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.
- Lievens, F., Conway, J. M., & De Corte, W. (2008). The relative importance of task, citizenship and counterproductive performance to job performance ratings: Do rater source and team-based culture matter? *Journal of Occupational and Organizational Psychology* , 81, 11-27.
- McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology* , 73, 327-330.
- Morrison, E. W. (1994). Role definitions and organizational citizenship behavior. *Academy of Management Review* , 37, 1543-1567.
- Motowidlo, S. J. (2000). Some basic issues related to contextual performance and organizational citizenship behavior in human resources management. *Human Resources Management Review* , 10, 115-126.
- 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.

- 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.
- Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance* , 2, 183-200.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal : Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage Publications.
- 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.
- Murphy, K. R., Herr, B. M., Lockhart, M. C., & Maguire, E. (1986). Evaluating the Performance of Paper People. *Journal of Applied Psychology* , 71, 654-661.
- Naylor, J. C., & Wherry, R. J. (1965). The use of simulated stimuli and the JAN technique to capture and cluster the policies of raters. *Educational and Psychological Measurement* , 25, 969-986.
- O'Brian, K. E., & Allen, T. D. (2008). The relative importance of correlates of organizational citizenship behavior and counterproductive work behavior using multiple sources of data. *Human Performance* , 21, 62-88.
- Organ, D. W. (1988). *Organisational citizenship behaviour: 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. (1988). *Organizational citizenship behavior: The good soldier syndrome*.  
Lexington, MA: Lexington Books.
- Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, trust, and organizational citizenship behavior. *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.
- Potter, E. H., & Fiedler, F. E. (1981). The utilization of staff member intelligence and experience under high and low stress. *Academy of Management Journal* , 24, 361-376.
- Quinones, M. A., Ford, K. J., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology* , 48, 887-910.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications, Inc.
- Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global aspects of job performance: A policy-capturing approach. *Journal of Applied Psychology* , 87, 66-80.



- Sackett, P. R., Berry, C. M., Wiemann, S. A., & Laczko, R. M. (2006). Citizenship and counterproductive behavior: clarifying relations between the two domains. *Human Performance* , 19, 441-464.
- Saegert, S. C., Swap, W., & Zajonc, R. B. (1973). Exposure, context, and interpersonal attraction. *Journal of Personality and Social Psychology* , 25, 234-242.
- Schmidt, F. L., Hunter, J. E., & Outbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology* , 71, 432-439.
- Schmidt, F. R., & Kaplan, L. B. (1971). Composite versus multiple criteria: A review and resolution of the controversy. *Personnel Psychology* , 24, 419-434.
- Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology* , 68, 653-663.
- Spiker, V. A., Harper, W. R., & Hayes, J. E. (1985). The effects of job experience on the maintenance proficiency of Army automotive mechanics. *Human Factors* , 27, 301-311.
- Van Scotter, J. R. (2000). Relationships of task performance and contextual performance with turnover, job satisfaction, and affective commitment. *Human Resource Management Review* , 10, 79-95.
- 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.

- Vey, M. A., & Campbell, J. P. (1994). In-role or extra-role organizational citizenship behavior: Which are we measuring? *Human Performance* , 17, 119-135.
- Wanous, J. P., & Hudy, M. J. (2001). Single-item reliability: A replication and extension. *Organizational Research Methods* , 4 (4), 361-375.
- Welbourne, T. M., Johnson, D. E., & Erez, A. (1998). The role based performance scale: Validity analysis of a theory-based measure. *Academy of Management Journal* , 41, 540-555.
- 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.
- Wilson, M. A., Johnson, E. C., & DuVernet, A. M. (2008). *Job analysis technical report*. Raleigh, NC: Mark Wilson Associates.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin* , 99, 432-442.
- Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research* , 17, 252-269.

## APPENDICES

## Appendix A.

### List of Trooper Job Tasks

Collision Investigation

Arrest & DWI Procedures

Court Preparation and Presentation

Equipment Use and Maintenance

Evidence Collection and Maintenance

Extraordinary Vehicle Operation

Officer and Public Safety

Preventative Patrol: Passenger Vehicles

Communication

Community Relations

Critical Incident Activities

Knowledge of Laws, Policies, and Procedures

Special Duty Activity

## Appendix B

### Contextual Performance Items

Subdimension		Item content
Altruism	A1	Helps others who have heavy workloads.
	A2	Is always ready to lend a helping hand to those around him/her.
	A3	Helps others who have been absent.
	A4	Willingly helps others who have work-related problems.
	A5	Helps orient new people even though it is not required.
Conscientiousness	CS1	Believes in giving an honest day's work for an honest day's pay.
	CS2	Attendance at work is above the norm.
	CS3	Does not take extra breaks.
	CS4	Obeys company rules and regulations even when no one is watching.
	CS5	Is one of my most conscientious employees.
Civic Virtue	CV1	Keeps abreast of changes in the organization.
	CV2	Attends meetings that are not mandatory, but are considered important.
	CV3	Attends functions that are not required, but help the company image.
	CV4	Reads and keeps up with organization announcements, memos, and so on.
Courtesy	CS1	Tries to avoid creating problems for co-workers.
	CS2	Considers the impact of his /her actions on co-workers.
	CS3	Takes steps to try to prevent problems with other workers.
	CS4	Does not abuse the rights of others.
	CS5	Is mindful of how his/her behavior affects other people's jobs.

## Appendix C

### Dissertation Proposal Document Introduction and Method Sections

#### A Multi-Level Investigation of Overall Job Performance Ratings

Job performance is typically conceptualized as “actions and behaviors that are under the control of the individual that contribute to the goals of the organization,” (Rotundo & Sackett, 2002, p. 66). Especially important to this definition are the ideas that performance is behavior (as opposed to the results of behavior) and that this behavior can be evaluated in terms of its contribution to organizational goals (Campbell, 1990). Campbell (1990) asserts that job performance is inherently multidimensional, a view that is shared by the vast majority of performance researchers (e.g., Borman & Motowidlo, 1993; Campbell, Gasser, & Oswald, 1996; Hesketh & Neal, 1999; Motowidlo, Borman, & Schmit, 1997; Murphy & Shiarella, 1997). However, it is also generally accepted that, while multidimensional conceptions of performance are appropriate when conducting research, for decision making in organizations, a unidimensional, or composite criteria is preferred (Schmidt & Kaplan, 1971). And while the notion that raters will differ in the way they combine information to arrive at an overall rating is far from new (e.g., Naylor & Wherry, 1965), in the job performance domain, research dedicated to explaining this variability is limited.

According to Campbell (1990), the process of assigning a rating to an individual’s job performance is as follows: observation, sampling, encoding, storage, retrieval, evaluation, differential weighting, and composite scoring. In the current study, I confine my focus to the stages of this process in which supervisors combine performance information in order to

assign an overall performance rating to their subordinate. Specifically, I am interested in whether or not different supervisors use different types of information to inform their ratings of subordinate overall performance and, if so, are there factors that can explain these differences? The notion that different individuals may arrive at the overall performance composite in different ways has some serious implications, both for research and practice. Murphy and Shiarella (1997) demonstrated that the validity of selection tests can vary considerably depending on how performance dimensions are weighted to form an overall performance composite. If individuals within an organization vary in the way they combine performance dimension information, the organization could have serious difficulties understanding or improving the performance of their selection system.

In addition to individual rater characteristics that may affect the way performance information is combined, situational or organizational factors (e.g., team-based culture; Lievens, Conway, & De Corte, 2008) may also play a role. If there is variance in the way overall performance composites are formed across organizations, this could mean serious problems for researchers seeking to draw conclusions about relationships between overall performance and other variables by generalizing across studies or organizations, as the label “overall performance” could be applied to very different constructs. Clearly, empirical research which investigates whether or not supervisors vary in how they judge subordinates’ overall performance is needed. On the practical side, research confirming that these differences exist would point to the value of rater training and/or the use of statistical procedures for combining dimension ratings. For researchers, an interesting next step might

be to identify whether there are factors related to the individuals or the work context that explain these differences. This would allow us to identify when performance ratings across raters or organizations are most likely to be comparable, and when they are not. In addition, by furthering our understanding as to how the work context influences the value of different types of behaviors, this research would benefit HR practices beyond performance appraisal, such as selection and training.

In the current study, I will investigate whether supervisors in a local law enforcement agency differ in the way they use performance dimension information when rating their subordinates' overall performance. Because I expect these differences may be related to characteristics of the rater as well as the individual subordinate and the environment in which the subordinate works, I adopt a multilevel approach. To this end, conducting this research in a law enforcement organization has a number of advantages. By using a sample of working adults, I can investigate performance rating in the context of real supervisors and subordinates who have developed real relationships over time. Moreover, compared to other organizations, law enforcement organizations are made up of large numbers of employees with identical job titles. Every employee in the proposed sample has the same formal job description and, as a result, differences in how certain performance behaviors are related to overall performance cannot be explained by differences in formal job duties. In addition, the organization is divided geographically into districts. This allows me to investigate the role of contextual factors related to the work environment. Many studies are not able to assess these types of factors as they are often shared across the entire sample.



In the following sections, I review prevailing theory related to the task and contextual dimensions of performance, as well as overall performance (or, more specifically, *ratings* of overall performance). The remainder of the introduction will describe my specific hypotheses and research questions as well as relevant empirical findings to date.

### *Task Performance*

Task performance refers to behaviors and activities that support the organization's 'technical core,' which can involve the execution of technical processes (transforming raw materials into the goods or services provided by the organization) or the maintenance of those processes, for instance by providing raw materials, distributing products, or through planning and coordination functions (Borman & Motowidlo, 1993; Motowidlo, et al., 1997). In a law enforcement context, activities that contribute directly to the organization's technical core would include issuing citations and making arrests, while activities such as hiring and training new officers serve to maintain the technical core. To perform job tasks successfully, the employee must have sufficient knowledge of both the task and the context in which it is performed as well as the ability to perform the task when appropriate (Murphy & Shiarella, 1997).

Two distinguishing features of task performance are that the requisite behaviors or activities are usually job-specific and that performance level tends to be predicted by the performer's level of knowledge, skill, and ability (Borman, 2006). Task activities are formally acknowledged by the organization as job requirements, often explicitly via job descriptions.

### *Contextual Performance*

*Construct definition.* The concept of organizational citizenship behavior (OCB; Smith, Organ, & Near, 1983) originally resulted from an interest in the behavioral consequences of job satisfaction (Motowidlo, 2000). Early measures of citizenship behavior were derived from interviews in which supervisors were asked to identify job behaviors that were “helpful, but not absolutely required,” (Smith, et al., 1983, p. 656).

Ten years after the introduction of OCB, a different line of research resulted in the contextual performance construct. Borman and Motowidlo (1993) believed that performance measures used in selection research and practice were deficient in that they focused solely on task performance, while ignoring activities such as persisting, helping, and endorsing organizational objectives, activities which the authors broadly termed, “contextual performance.” The authors argued that the criterion domain consists of both task performance, behaviors that contribute to the organization’s technical processes, and contextual performance, or behaviors that support the broader psychological and social environment in which that technical core must function (Borman & Motowidlo, 1993). Unlike Smith and colleagues (1983), who confined their OCB construct to those behaviors that are “not absolutely required,” Borman and Motowidlo made no such restriction regarding contextual performance. Rather, for a behavior (required or otherwise) to be considered contextual performance, it must contribute to organizational goals *indirectly*, through the work context.

Despite the difference in definition, the two constructs were found to share many of the same behavioral elements, a fact that quickly became a source of confusion for researchers (Motowidlo, 2000). Perhaps in response to this concern, Organ (1997) published an article in which he redefined the OCB construct to be synonymous with contextual performance, noting that the earlier requirement that OCBs be extra-role was problematic given that roles tend to be fuzzy and changeable, and are often negotiated between supervisor and subordinate.

Rather than recommend that OCB be abandoned altogether, though, Organ proposes a solution that amounts to changing the name of the contextual performance to OCB. The problem with the term contextual performance, he argues, is that is it, “cold, gray, and bloodless,” (Organ, 1997, p. 91). However, because the definition of contextual performance has been more consistent over the years and because I believe the term better represents the defining feature of this performance dimension (i.e., that behavior contributes to organizational effectiveness through the work *context*), I adopt this terminology for the remainder of this paper.

*Dimensionality.* In their initial study, Smith and her colleagues (1983) specified only two classes of OCB, Altruism (helping behaviors directed at other individuals) and Generalized Compliance (behaviors directed toward the system). Later, Organ (1988) revised the construct to include five facets, Altruism, Conscientiousness, Sportsmanship, Courtesy, and Civic Virtue. This five category model has gained considerable visibility in the literature, possibly due in part to the popularity of Podsakoff, MacKenzie, Moorman, and Fetter’s

(1990) measure based on this structure. However, Organ (1997) notes that Altruism and Courtesy items often fail to load on their intended factors. Organ further suggests that citizenship behaviors might be more usefully classified using a two dimensional framework like the one developed by Williams and Anderson (1991). The Williams and Anderson (1991) model specifies that the altruism and courtesy dimensions represent OCB directed towards individuals (i.e., OCB-I), while the conscientiousness, sportsmanship, and civic virtue dimensions comprise OCB directed toward the organization (i.e., OCB-O).

A similar model emerged from Coleman and Borman's (2000) rational approach to clarifying the contextual performance construct domain. In their study, 44 industrial and organizational psychologists were asked to sort behaviors drawn from existing models of citizenship and contextual performance based on their similarity. The resulting model specified three broad categories of behavior: interpersonal citizenship performance, organizational citizenship performance, and job-task conscientiousness.<sup>1</sup> The first two categories parallel those defined by Williams and Anderson (1991), while the third category captures behaviors that benefit the job or task.

---

<sup>1</sup> While the model has been well-received, it should be noted that other research has suggested that the job-task conscientiousness dimension might be more appropriately viewed as representing both task and contextual performance (J. W. Johnson, 2001; Motowidlo, et al., 1997). An earlier study in which Van Scotter and Motowidlo (1996) found a similar dimension, job dedication, to be indistinguishable from task performance, would seem to support this argument.

Still other investigations, however, have failed to provide support for multidimensional conceptualizations of the contextual performance domain. At least two recent meta-analyses (i.e., Hoffman, Blair, Meriac, & Woehr, 2007; LePine, Erez, & D. E. Johnson, 2002) support a single factor model. Nonetheless, an examination of recently published studies reveals that many researchers still differentiate between individually- and organizationally-directed behaviors (e.g., Ilies, Nahrgang, & Morgeson, 2007; Kamdar, McAllister, & Turban, 2006; O'Brian & Allen, 2008). Moreover, the results of these studies seem to suggest that examining the two classes of behavior separately may improve our understanding of some organizational processes. For example, the Ilies study found that the quality of relationships between leaders and followers is more strongly related to individually-targeted behavior than to organizationally-targeted behavior. Similarly, the Kamdar study concludes that interpersonally-directed behaviors seem to be more strongly related to relational variables, while organizationally-directed behaviors are more the result of exchange-related variables. Taken together, the results of these studies indicate that the two types of contextual performance may likely be the result of distinct processes. Still others have found the Coleman and Borman 3-factor model to provide a better fit to their data than a single-factor model (e.g., Sackett, Berry, Wiemann, & Laczko, 2006). For now, it seems that the dimensionality of contextual performance is far from settled. In the current study, I operationalize contextual performance in terms of both individually- and organizationally-directed behaviors. Based on the research discussed above, I believe that the potential for

differential relationships is worth exploring, although I do not make explicit hypotheses regarding these differences.

### *Overall Performance Ratings*

An individual's overall performance rating can be thought of as a measure of his or her organizational worth (Motowidlo & Van Scotter, 1994). This is because overall job performance represents the aggregate contribution of an employee's behaviors to the organizational goals, over the course of the evaluation period (Motowidlo, et al., 1997). Implicit in this definition is that job performance must include *only* those behaviors that contribute to the organization's goals (Campbell, 1990). But, which behaviors are these? Campbell explains that there are two judgments involved in this determination. The first is a *value* judgment as to what the organizational goals actually are. Sometimes organizational goals are formally articulated, but often there is some ambiguity as to what really matters. The second judgment is an *expert* judgment and refers to the decision as to whether or not observed behaviors contribute to the organization's goals.

The model seems to suggest that the rating assigned to an incumbent's overall job performance will depend not only on the level at which he or she performs certain behaviors, but also on the rater's beliefs about the goals of the organization and his or her mental model relating behaviors to goals. If a rater believes that a certain behavior contributes very little to the organizational goals he or she deems valuable, the behavior should have very little influence on the overall rating, even if the incumbent performed exceptionally well in that area.

Related to this is the idea that supervisors hold certain ‘role expectations’ regarding what their employees are supposed to do. According to role theory, the role expectations that one holds are influenced by characteristics of both the individual and the context in which he or she is embedded (Welbourne, D. E. Johnson, & Erez, 1998). Lievens et al. (2008) suggest that performance evaluation represents an appraisal of how well a worker’s behavior conforms to the role expectations held by the rater (typically a supervisor). Thus, if a worker’s supervisor does not believe that contextual performance behaviors are an important aspect of that worker’s job, those behaviors will not weigh heavily into that supervisor’s assessment.

In the current study, I propose a hierarchical investigation of the influence of various factors on ratings of overall job performance. By using a hierarchical linear modeling (HLM) approach, it is possible to investigate factors that are specific to an individual employee, such as his or her time on the job, as well as factors that are shared among employees, such as characteristics of the supervisor or of the work environment. In particular, I expect that many of these factors will play a role in whether or not supervisors will view contextual performance as role-required and, as such, will affect its weight in supervisors’ evaluations of their employees’ performance. In the following sections, I discuss these factors and offer hypotheses and questions for research. For clarity, I begin with a discussion of relationships at the lowest level (Level 1), or those that involve factors that vary among individual employees, and then turn to higher-level (Level 2) models, where I discuss the possible influence of those factors which vary at the supervisor-level.

### *Level 1 Models*

That an employee's overall performance rating will be related to both the task and contextual components of his or her work performance is a fairly well-established finding. For instance, in their (1994) study involving US Air Force mechanics, Motowidlo and Van Scotter found that task performance explained between 17 and 44% of variance in overall performance ratings assigned to the mechanics, while contextual performance explained an additional 12 to 34% of overall performance rating variability, depending on the particular reliability corrections employed. Importantly, in their study, separate raters were assigned to rate the different aspects of performance (i.e., task, contextual, and overall), thus limiting the influence of rater effects on the observed results. As a result, their study was not able to determine how much variability existed within individual raters with respect to weighting the task and contextual performance factors.

Podsakoff, MacKenzie, Paine, and Bachrach (2000) provide an in-depth review of the research in this area, in which they discuss similarities and differences among the relevant behavioral constructs proposed in the literature (e.g., OCB, contextual performance, etc.), the proposed antecedents and consequences of these behaviors, and, finally, the future directions that they perceive to be most promising. For the current discussion, however, perhaps the most relevant conclusion drawn by the authors is that the independent contributions of task and contextual performance to overall performance have been supported not only in field studies, but also experimental and meta-analytic investigations.



A more recent review (Bergeron, 2007) lists the findings of published studies regarding the weight of task and contextual performance in predicting a number of work outcomes, including performance evaluations, promotions, and other rewards. This review also supports the conclusion that both aspects of performance have been found to contribute to overall performance evaluations across studies, although Bergeron also notes that, in the majority of these studies, task performance tended to be weighted more (i.e., of the 19 samples included in the list, task performance was given more weight in 12, contextual performance was given more weight than task performance in 2, while 5 indicated mixed results).

Based on the substantial empirical evidence that task performance and contextual performance are distinct constructs (e.g., LePine, et al., 2002), both of which contribute to ratings of overall performance (e.g., Borman, White, & Dorsey, 1995; Conway, 1999; Motowidlo & Van Scotter, 1994), I predict:

*Hypothesis 1:* Ratings of both task performance and contextual performance will be positively related to overall performance ratings.

In addition, because Bergeron's (2007) review of published studies indicates that task performance is usually, but not always, given more weight in overall performance ratings, I will also compare the variance in overall performance ratings that can be attributed to task and contextual performance ratings:

*Research Question 1:* What proportion of the variance in overall job performance ratings can be attributed to task vs. contextual performance?

One limitation shared by the studies described above is that they ignore the possibility that raters have different ways of using information about task and contextual performance when assigning overall performance ratings. That is, while the above studies have demonstrated that, across raters, ratings of both task and contextual performance are related to overall performance ratings, few attempts have been made to determine whether these relationships vary, perhaps as a function of the rater or situation.

In efforts to overcome this limitation, some more recent studies have adopted an experimental policy-capturing methodology. According to Rotundo and Sackett (2002), typical policy capturing studies consist of, “(a) presenting raters with a series of profiles in which the independent variables of interest are manipulated, (b) obtaining raters’ judgments about the dependent variable, and (c) using multiple regression analyses to compute the relative importance of each independent variable,” (p. 70). In performance rating studies, these profiles would describe the performance of a hypothetical employee (i.e., a paper person), with levels of task and contextual performance manipulated. The performance profiles are typically designed so that dimension scores are uncorrelated. This property allows the relative importance of the performance dimensions to be calculated by regressing overall performance ratings on the dimension scores (J. W. Johnson, 2001). The result is a “captured rating policy” equation for each rater, which describes the way the rater weights performance dimensions in his or her assessment of overall performance (J. W. Johnson, 2001; Rotundo & Sackett, 2002).

Because policy-capturing studies result in a separate equation for each rater, they overcome some of the limitations associated with other performance appraisal studies, by providing more in-depth information about the differences between individuals (Karren & Barringer, 2002). However, the experimental nature of policy-capturing studies presents a different set of limitations. One might question whether the captured rating policies are valid indicators of the rating policy a supervisor would utilize in the work setting. Rating the described behavior of a paper person clearly involves a fundamentally different cognitive process compared to the task of rating actual, observed performance of real-world employees (Murphy, Herr, Lockhart, & Maguire, 1986). Other criticisms related to the use of paper people for performance research point out that performance profiles tend to contain less irrelevant information and that performance information is less ambiguous than is the case with real-world observations of performance (Murphy, et al., 1986). Additionally, the common practice of using uncorrelated dimension scores is unrealistic, as performance dimensions would rarely be expected to be uncorrelated in the real world (J. W. Johnson, 2001). Laboratory studies also fail to account for the on-going relationship between supervisor and employee or other potentially influential aspects of the political, social, and affective context in which performance appraisal occurs (Bretz, Milkovich, & Read, 1992).

Despite these limitations, the results of these studies suggest that raters do vary in the way they combine information about performance dimensions into ratings of overall performance. Specifically, using the policy capturing approach, Rotundo and Sackett (2002) demonstrated that raters do not share a common policy for weighting performance

dimensions when they evaluate overall performance. Building on this research, Lievens and colleagues (2008) found that their organizational culture manipulation (high vs. low team-based culture) accounted for significant observed variance in raters' policies for combining performance dimensions. It seems reasonable to expect that if raters differ in the way they use written performance information in the laboratory, the same may be true in the field. However, research is needed to determine whether or not these findings do generalize to the field and, if so, to investigate what other factors may have an influence.

In the present study, I propose to use hierarchical linear modeling (HLM) to investigate whether various characteristics of the subordinate (ratee), supervisor (rater), and work environment influence the relationships between performance dimensions and overall performance ratings. As with the policy-capturing approach, each supervisor will rate multiple employees. Also similar to the policy-capturing approach, the HLM approach involves the estimation of a separate equation for each rater/supervisor. However, unlike the policy capturing approach, the proposed methodology does not rely on experimental methodology, but rather involves real employees in a real work environment.

Having already predicted that both task and contextual performance will relate to overall performance, the next question is whether these relationships will vary across supervisors. As the above-described policy capturing studies have concluded that, "global ratings of performance rely to a great extent on raters' implicit policies of what really matters," (Rotundo & Sackett, 2002, p. 78), there is reason to believe that at least some will.

However, because the current study examines performance ratings of real employees, I expect that both the organization's and the supervisor's policy will play a role.

A number of authors (e.g., Bergeron, 2007; Murphy & Cleveland, 1995; Organ, 1997) have suggested that, while not always the case,<sup>2</sup> it is often useful to think of task performance as those behaviors and activities listed in an employee's formal job description. When characteristics of effective task performance are formally prescribed by the organization, as is the case with formal job descriptions, one might assume that the importance of such behaviors will be understood by all relevant organizational members. In this case, these activities and behaviors are publically recognized by the organization as defining aspects of the job, or what the employee has been hired to do and will be evaluated for doing. Accordingly, I argue that it is the organization's policy, rather than the individual supervisor's policy, that dictates the relationship between an employee's task performance and his or her overall job performance.

*Hypothesis 2(a):* The strength of the relationship between task performance and overall performance will not vary across supervisors.

Compared to task performance, the organizational policy regarding contextual performance is likely less clear. Even though the prevailing definition does not limit contextual performance to discretionary behaviors, contextual behaviors are less likely than

---

<sup>2</sup> Of course, this is more likely to be the case when organizations or jobs are not undergoing serious change, and when job descriptions are the result of recent and comprehensive job analyses. In the proposed study, both of these conditions are true.

task behaviors to be formally required or clearly linked to rewards (Organ, 1997). As the contribution of contextual performance behaviors to overall performance is not generally expected to be formalized by the organization, the possibility exists that, within an organization, individuals will vary in their beliefs about contextual performance. In support of this possibility, Morrison (1994) found that employees' beliefs about whether or not various contextual performance items were part of their job role varied within a single organization. If employees differ in the extent to which they see contextual behaviors as part of the job they are hired to do, supervisors might also. Unlike the task performance-overall performance relationship, which I believe to be dictated by the organization and therefore predict will not vary across supervisors, I believe the link between contextual performance and overall performance is largely determined by the implicit policy of the individual. If this is the case, the slope relating subordinates' contextual performance to overall performance will vary across the supervisors providing those ratings. In other words:

*Hypothesis 2(b):* The strength of the relationship between contextual performance and overall performance ratings will vary across supervisors.

*Other level-1 influences.* Another employee-level variable that might influence ratings of overall performance is work experience. Work experience has been defined as the events an employee perceives or experiences that relate to the performance of his or her job (Quinones, Ford, & Teachout, 1995). In practice, work experience has been operationalized as time on the job (e.g., Schmidt, Hunter, & Outbridge, 1986), time in the organization (e.g., Potter & Fiedler, 1981), time in the occupation (e.g., McDaniel, Schmidt, & Hunter, 1988),

and number of times performing a given task (e.g., Spiker, Harper, & Hayes, 1985), to name a few. Quinones and colleagues' (1995) meta-analysis demonstrated that, regardless of the operationalization used, work experience tends to be positively related to overall job performance (estimated  $\rho=.27$ ). Their results suggest that even though tenure-based measures of experience, such as time on the job, clearly do not capture the richness of an employee's specific work experiences, it is still reasonable to expect that such a measure would significantly correlate with measures of overall job performance. Additional support for this position is provided by an earlier meta-analysis which statistically linked years on the job with performance across a range of jobs (although the relationship was stronger for those jobs with lower cognitive demands; McDaniel, et al., 1988). Based on the results of these studies, I predict:

*Hypothesis 3:* Employee (trooper) time on the job will be positively related to overall performance ratings.

Further, in his call for research which addresses the impact of employee characteristics on the relative importance of task and contextual performance, J. W. Johnson (2001) suggests that an employee's level of experience may influence how his or her supervisor weights the dimensions. I suspect that, at least in some organizations, employee work experience moderates the influence of contextual performance on assessments of overall performance. Organ (1988) for instance, theorized that, as individuals move up the organizational ranks, the breadth of their roles expands. Likewise, in her recent review, Bergeron (2007) notes that in studies based on samples of salespeople, task performance

tends to be given greater weight in performance evaluation than contextual performance, whereas the same does not hold true for studies which involving higher ranking employees, such as managers. Rather, in managerial studies, contextual performance tends receive equal, and sometimes greater weight than task performance.

Perhaps employee tenure within ranks has a similar effect. That is, when an employee is relatively new to an organization and still ‘learning the ropes,’ his or her expected contribution to the organizational goals might be restricted to those task activities formally prescribed by the organization (i.e., those activities he or she learned in training). As the employee’s skills and abilities related to performing these tasks improve, so does task performance and, likewise, his or her contribution to the organization. Eventually, however, one would expect that employees will either master these skills or leave the organization. For employees who master the work tasks, it is conceivable that role expectations may expand. An experienced employee’s failure to perform task activities when needed would still be seen as detrimental to the effectiveness of the organization and exceptionally high levels of task performance would still be seen as valuable, just as one might expect for lower tenure employees. However, because the more experienced employee has mastered those tasks, the resources he or she must expend to perform them effectively ought to be less than those required for an inexperienced employee to perform the same tasks (e.g., Ackerman, 1986, 1987; Murphy, 1989). Thus the experienced employee might be expected to contribute to the organization’s goals in additional ways, perhaps by assisting or encouraging inexperienced employees, or otherwise enabling the effective execution of task activities.



I predict that role expansion will occur as employee tenure increases, such that the overall performance score assigned to an employee will correlate more strongly with contextual performance as that employee's experience increases. In addition to the intuitive appeal of the above described process (which, to my knowledge, has not been empirically tested), because the highway patrol's model of career advancement centers around internal promotion, it may be the case that high tenure employees are expected to take on expanded job roles in anticipation of a promotion to a higher rank.

*Hypothesis 4:* Employee (trooper) time on the job will moderate the relationship between contextual performance and overall performance ratings, such that the relationship is stronger for those with more (vs. less) time on the job.

Besides characteristics of the individual employees, factors specific to the individual employee's relationship with his or her supervisor can be included among Level 1 variables. In the current study, I include one such factor, namely the supervisor's opportunity to observe the employee's performance.

Judge and Ferris (1993) argue that, due to prevalence of factors motivating supervisors not to assign low ratings to performance (e.g., lawsuits, paperwork associated with complaints, resentment), the performance rating process is akin to a search for positive information. In this case, increased opportunity to observe would increase the body of information to search, increasing the likelihood that the supervisor will succeed in finding justification to assign high rating. While I should note that the Judge and Ferris study did empirically support the positive link between opportunity to observe and overall performance

ratings, one might also argue that, if an employee truly is a poor performer, increased opportunity to observe should increase the likelihood that the supervisor will detect this deficiency. Importantly, the current study proposes to investigate job performance in a law enforcement context. Presumably, in high stakes jobs a supervisor must weigh the advantages of inflating performance ratings against the potential risks of doing so. While it may be tempting to inflate an employee's performance rating to avoid a complaint, lawsuit, or the like, the supervisor must also consider whether failure to document performance deficiencies could result in public or officer safety issues, which might also incur complaints or lawsuits. Thus, in the law enforcement, supervisors may be searching for both positive and negative information about the employee's performance. As such, it is unclear what (if any) effect opportunity to observe will have on overall performance.

*Research Question 2:* Is opportunity to observe related to overall performance ratings?

#### *Level 2 Models*

Among the supervisor (rater) characteristics that might affect performance ratings is supervisor experience. Research findings regarding this relationship have been mixed, with some studies reporting supervisors with more tenure tend to be more lenient, and others finding no effect (Landy & Farr, 1980). In their 1993 study, Judge and Ferris offer a number of plausible reasons why one would expect supervisor job experience to relate positively to ratings of overall performance. They speculate that with increased tenure comes increased confidence and decreased need to prove one's toughness via harsh ratings. Experienced

supervisors may also have come to understand that the relationship between behavior and performance is a tenuous one and may have a better grasp on the potential costs of assigning unfavorable ratings (e.g., additional paperwork associated with complaints, employee resentment or hostility). However, their study failed to find evidence of a link between supervisor experience and overall performance ratings. Thus, in this study, I include the role of supervisor experience as a research question, rather than a hypothesis:

*Research Question 3:* Will supervisor experience (time in current rank) positively influence overall performance ratings?

Perhaps the more interesting question regarding supervisor experience, though, is whether it might moderate the relationship between ratings of contextual performance and overall job performance. That is, as one gains experience supervising a certain job, does he or she come to see the contribution of contextual performance behaviors to the overall goals of the organization differently? If so, does the relationship become stronger or weaker?

Because the sample proposed for the current study comes from an organization where internal promotion is the norm, one might argue that supervisors with more experience in their current rank will have had time to internalize their role as a supervisor, while those with less experience might still be transitioning from their previous role. These less experienced supervisors may well be rating the performance of employees with whom, until recently, they worked alongside as a peer. If this is the case, previous research would seem to suggest that the relationship between contextual performance and overall performance ratings will be stronger when those ratings are provided by less experienced supervisors. For example, in

their recent study, Lievens and colleagues (2008) compared the relative weights associated with task and contextual performance dimensions across rating sources and found that peers tend to give more weight to contextual performance in their assessments of overall performance than do supervisors. Similarly, Conway's (1999) meta-analysis demonstrated that peers attend more to interpersonal facilitation when assessing overall performance than do supervisors.

Vey and Campbell (1994) conducted a study in which undergraduate students were asked to identify various work behaviors as in-role or extra-role. The authors found the supervisory experience of the students to be negatively related to the tendency to see the altruism, courtesy, and sportsmanship facets of organizational citizenship behavior as reflecting in-role behaviors. The authors point out that, as traditional measures of organizational citizenship were developed by asking managers to enumerate behaviors that they considered to be "extra-role," it is not surprising that supervisory experience might be negatively related to perceptions of such behaviors as "in-role." Further, it is often peers who stand to benefit the most from behaviors such as altruism or courtesy, whereas supervisors' relative removal from the day to day work activities may mean that they are less likely to see how these behaviors contribute to effectiveness. In support of the idea that interpersonal behaviors are more relevant to peers than they are to supervisors, Borman et al. (1995) found that the perceived 'obnoxiousness' of a ratee figured into appraisals made by peers far more than it did in appraisals made by supervisors.

Alternatively, it is also plausible that relatively inexperienced supervisors may experience insecurity about their new role and, as a result, focus their performance management efforts around those behaviors and activities that are formally prescribed by the organization (i.e., task performance), to the exclusion of contextual performance. Along these lines, Befort and Hattrup (2003) argue that novice supervisors will tend to focus on concrete tasks and problems more so than experienced supervisors who, through their experience, better understand the importance of the psychological and social context of work. The authors point out that, because less experienced supervisors themselves tend to be supervised more closely than more experienced supervisors, it would typically be in their best interest focus their attention on the task activities of their subordinates, which presumably are more observable and more directly linked to the goals of the organization.

Because it seems that compelling arguments can be made for a positive or negative cross-level moderation effect, I propose to investigate the role of supervisor experience as a research question, rather than a hypothesis:

*Research Question 4:* Does the relationship between contextual and overall performance ratings depend on how long the supervisor providing those ratings has held the rank of sergeant?

As I have discussed previously, when performance is evaluated as the contribution of an individual's behaviors to the goals of the organization, the value of those goals must be judged by the organization or, more realistically, certain organizational members (Campbell, 1990). While Campbell does not speculate as to what types of information might form the

basis of this value judgment, characteristics of the environment in which the organization operates would seem relevant. Murphy and Cleveland (1995) hypothesize that characteristics of the environment likely affect performance appraisal through a number of mediating variables, including standards for performance and performance dimensions. Regarding performance dimensions, they write, “the definition of what constitutes performance and the relative importance of different parts of the performance domain are likely to be affected by the environment,” (p. 39). Unfortunately, as the authors go on to note, research addressing the effects of the environment on performance appraisal is “virtually nonexistent,” (p. 43). More recently, Hanson and Borman (2006) lament that “little is known about the mechanisms by which citizenship performance affects organizational effectiveness or supervisor ratings,” but they tentatively suggest that job or situational characteristics may moderate the relationship (p. 166).

Returning to the distinction between the task and contextual dimensions of job performance, it is generally believed that task activities are performed because they are explicitly required by the job, while contextual activities tend to have a more to do with an employee’s motivation. However, the assumption that any individual who is motivated to engage in contextual performance behaviors will do so may not always be accurate (Griffin, Neal, & Neale, 2000). Rather, aspects of the work situation, such as task demands, may affect opportunities to perform the more discretionary aspects of the job. The authors believe that, when task demands are high, the majority of an individual’s time and energy will be consumed by performing task activities, leaving few resources available for contextual

behaviors. As Bergeron (2007) puts it, “Recognizing that individuals have limited time, any time spent on OCB comes at the expense of task performance,” (p. 1078).

Along these lines, Smith and colleagues’ (1983) seminal study found that that individuals with rural (vs. urban) backgrounds are more likely to engage in altruistic behaviors. The authors explain that, “people from urban settings learn to cope with overstimulation from the physical and social environment by screening out many social stimuli, thus decreasing their sensitivity to others' needs,” (p. 662). The current study proposes to incorporate this notion of urban versus rural work setting by including two variables expected to shape the work environment, namely, population and number of interstate highways. An environment characterized by more highways and a higher population should mean increased stimulation from the physical and social environments, and, accordingly, increased task demands. In the Griffin study, the authors found that for a sample of air traffic controllers, supervisors’ ratings of subordinate task performance were related to ratings of overall effectiveness across situations, while contextual performance (teamwork, professionalism, and supporting organizational objectives) was only related to effectiveness in situations in which task demands were low (based on variables such as traffic volume and traffic complexity). Moreover, if working in an urban district causes supervisors to screen out interpersonal stimuli, as Smith and colleagues suggest, urban supervisors would probably assign less importance to the social context of work and, therefore, struggle to see the value of contextual performance. Consequently, I expect supervisors’ ratings of their

subordinates' contextual performance will be more strongly related to overall performance ratings when district population and number of highways are low:

*Hypothesis 5:* The relationship between contextual performance and overall performance ratings will be stronger (vs. weaker) for districts that contain a lower (vs. higher) number of interstates.

*Hypothesis 6:* The relationship between contextual performance and overall performance ratings will be stronger (vs. weaker) for districts with lower (vs. higher) population.

In sum, many questions remain with respect to the variance in raters' overall performance ratings that may be due to task and contextual performance. In the following section, I discuss the proposed sample, measures, and analytical procedure used to test these hypotheses and research questions.

## Method

### *Participants*

The proposed sample consists of North Carolina State Highway Patrol (NCSHP) troopers and their supervising sergeants selected to participate in a recent selection system revalidation effort (see E. C. Johnson, DuVernet, & Wilson, 2008). While the original validation sample consisted of 447 troopers (33.4% of the total trooper population), this sample will be reduced somewhat due to the nature of the proposed hypotheses and research questions. First, because performance ratings are of central interest, only troopers whose supervising sergeants provided usable performance ratings will be included. In addition, the



hierarchical nature of the analyses (discussed below), require that each included sergeant provide two or more usable ratings. Thus, if a trooper was rated by a sergeant who did not rate any other troopers, that trooper's ratings cannot be included in the analyses. The following sections describe the reduced samples of troopers and sergeants.

*Supervisors.* As mentioned above, only those supervising sergeants who provided usable performance ratings for at least two troopers will be included in the analyses. Among those sergeants ( $N=90$ ), the number of troopers rated ranges from 2 to 13, with a median = 3 ( $M = 3.72$ ,  $SD = 2.07$ ). The majority reported that, at the time of rating, they had been in the role of sergeant for over three years. Unfortunately, no other demographic information is available for this sample.<sup>3</sup> The complete breakdown of job tenure for the sergeants who provided ratings can be found in Table D.1.

*Employees.* Usable performance ratings are available for  $N=335$  troopers. Time between being sworn in as a NCSHP trooper and the date of the evaluation ranges from 1.96 years to 7.81 years ( $M = 4.85$ ,  $SD = 1.85$ ). Table D.2 summarizes the demographic characteristics of the sample.

### *Procedure*

---

<sup>3</sup> While no demographic information is available for the supervisor sample ( $N=90$ ), there is no reason to expect that they differ substantially from the total sergeant population ( $N=222$ ), which is 16.2% African American, 1.8% American Indian, 81.5% Caucasian, and 96.4% male.

All performance measures were collected online as part of a validation effort for a selection system (see E. C. Johnson, et al., 2008). Sergeants responsible for supervising the troopers in the sample received a memo which directed them to the rating website and provided them with a list of the names and identification codes for the troopers they were assigned to rate. Sergeants who had supervised a trooper for fewer than six months were instructed to contact the trooper's previous supervisor and become familiar with the trooper's previous performance documentation before completing the ratings.

The online rating website instructed raters that all ratings were being collected for research purposes only and that no ratings would be seen by any member of the NCSHP. Raters entered a "trooper identification code" to indicate the particular trooper they were rating, and then made ratings on 46 items pertaining to the trooper's current levels of task performance, contextual performance, and overall performance.

### *Measures*

*Task performance.* Supervisors rated troopers' performance on 16 job tasks derived from a recent comprehensive job analysis conducted for the job title of Trooper (see Wilson, E. C. Johnson, & DuVernet, 2008). During that process, troopers from throughout the state were interviewed and observed doing their jobs by trained job analysts. The 16 job tasks were generated by the same job analysts in collaboration with a project advisory board of troopers, sergeants, and higher ranking NCSHP members.

Supervisors rated troopers' performance on each of the 16 tasks, using a 7-point scale, which ranged from "poor" to "excellent." The tasks are listed in Appendix A. Detailed

descriptions of each of the individual tasks were provided on the rating website. Only those 13 tasks that apply to all troopers will be included in the analyses (i.e., all tasks except for Inspection, Preventative Patrol: Commercial Vehicles, and Training Employees). Task Performance scores will be computed as the average rating across the remaining 13 tasks ( $\alpha=.93$ ). Principle components analysis indicated that a single factor (Eigenvalue=7.1) explained 54.9% of the total variation in the task performance data. A second factor (Eigenvalue=1.0) added only 7.7% to the cumulative explained variance. Note that while the eigenvalue > 1 test (Kaiser, 1960) could be used to defend a 2-factor solution, other authors have pointed out that this test tends to result in the over-extraction of factors, as compared to other alternative procedures, such as the scree test (e.g., Cortina, 2002; Zwick & Velicer, 1982; 1986).

Based on recommendations that predictor variables be centered before being entered into hierarchical analyses, each trooper's raw task performance score will be centered using the grand-mean centering approach, wherein the mean score of the entire sample is subtracted from the individual's score. Centering eases interpretation of model coefficients (i.e., in a model in which task performance predicts overall performance, the intercept would be interpreted as the overall performance rating for a trooper with average task performance) and may also improve model estimation (Kreft, de Leeuw, & Aiken, 1995).

*Contextual performance.* In addition to rating job task performance, supervisors completed an online version of a 24-item scale (Podsakoff, et al., 1990) designed to assess the five dimensions proposed by Organ (1988): Altruism, Conscientiousness, Civic Virtue,

Courtesy, and Sportsmanship. Supervisors indicated the degree to which they believed statements were true of the trooper they were evaluating, using a 7-point scale, which ranged from “strongly agree” to “strongly disagree.” The scale items can be found in Appendix B.

Contextual performance scores will be centered using the same grand-mean centering technique described for task performance. Separate scores will be computed for contextual performance directed towards the organization and interpersonally directed contextual performance. Based on the results of exploratory factor analysis (EFA) using the principle factors method and Promax rotation (Table D.3, interpersonally-targeted contextual behavior will be represented by averaging across the 5 altruism and 5 courtesy items ( $\alpha=.93$ ) while organizationally targeted behaviors will be represented by the 5 conscientiousness and 4 civic virtue items ( $\alpha=.86$ ). This is slightly different from the mapping proposed by Williams and Anderson (1991), which included sportsmanship items as indicators of organizationally-targeted behaviors. However, the EFA results indicated the items seem to form a factor unrelated to either interpersonal or organizational contextual performance, perhaps due to their negative wording. Including these items would reduce the reliability of this scale to .44. As a result, they are not included in the proposed analyses.

*Overall performance.* After responding to the task performance items but before the contextual performance items, supervisors also provided ratings of troopers’ overall performance, by responding to a single item, “After considering everything you know about this Trooper, how would you rate his/her overall performance?” Ratings were made using a

7-point scale, where 1 indicates “poor” and 7 indicates “excellent.” The actual ratings provided ranged from 2 to 7 ( $M = 4.99$ ,  $SD = .70$ ).

### *Performance Context*

The NCSHP is organized geographically into 8 troops, which are further divided into districts. While higher ranking members of the patrol may be responsible for multiple districts, troopers and sergeants are typically assigned to work in one specific district, which can be made up of 1 to 5 counties. In the current study, usable performance ratings are available from supervisors in 62 different districts, and work context will be described as the population and number of highways for the particular district in which a supervisor (and his or her employees) work. While it is possible to conceptualize the work environment as a third level of analysis (i.e., employees are nested within supervisors, which are in turn nested within districts), taking this approach would mean that only those districts from which multiple supervisors provided ratings. Because there are only 21 districts that meet this requirement, there simply would not be enough power to detect the cross-level effects, should they exist. For example, Hofmann (1997) asserts that to have adequate power to detect cross-level effects in samples that contain 30 higher-level units, one would need to have at least 30 lower-level observations nested within each of those units. Because the number of supervisors providing ratings within each district is nowhere near 30 (in fact, it seldom exceeds 2 and never exceeds 3), in the current study, work context will be modeled as a level-2 variable and data from all 62 districts will be included.

### *Other Measures*

In addition to making the above described performance assessments, the supervising sergeants also provided information about how long they had supervised the trooper they were rating (*“How long have you supervised this Trooper?”*) and how often they have the opportunity to observe the trooper’s performance (*“How often do you get to directly observe this Trooper doing their job?”*). Approximately half of the supervisors reported that they had supervised the trooper for at least 2 years (see Table D.4 for a complete breakdown) and two thirds reported observing the trooper’s performance weekly (see Table D.5).

*Proposed Analysis: Hierarchical Linear Modeling (HLM)*

The variables described above can be summarized as characteristics of troopers, characteristics of the sergeants who supervise them, and characteristics of the work environment. Importantly, these variables exist at multiple levels of analysis: there are variables describing the troopers, and those troopers are grouped within the sergeants who supervise them and evaluate their performance.

There are three main approaches for analyzing multilevel data (Hofmann, 1997). The first option is to disaggregate the data. In the present discussion, this would mean that each trooper would be assigned scores representing characteristics of their supervisor and work environment and all analyses would be conducted at the trooper level. Unfortunately, there are two problems associated with this approach. The first is that ratings of different troopers completed by the same supervisor are likely to be more similar than ratings of different troopers completed by different supervisors. In this way, nested data violate the assumption of independent observations inherent in regression analyses (Raudenbush & Bryk, 2002).

Second, all tests would be based on the number of individual ratings ( $N=335$ ), while the number of supervisors actually making those ratings ( $N=90$ ) is markedly smaller. As a result, calculated standard errors would likely be underestimates and significance tests would likely be too lenient (Raudenbush & Bryk, 2002).

A second option would be to create an aggregate rating for each sergeant and conduct all the analyses at the sergeant level. This is clearly not ideal, however, as this approach would, in effect, throw out all trooper-level information such as individual differences in performance.

The third approach is to analyze the data via hierarchical linear models which take into account the multilevel nature of the data and allows group and individual level variables to be investigated at the appropriate levels of analyses (Hofmann, 1997).

Hierarchical linear models allow that observations drawn from the same group may not be independent. For example, troopers who report to the same sergeant likely share values on unmeasured variables such as goals set by their supervising trooper. Using traditional approaches, such as ordinary least squares (OLS) regression, these variables would be treated as error, resulting in correlated residuals (Raudenbush & Bryk, 2002). HLM, however, allows group- and individual-level residuals to be modeled separately, recognizing that observations may be dependent. Because ratings of different troopers made by the same supervisor are likely not completely independent, this approach represents the most appropriate analytic tool for the current proposal. The use of HLM will also allow the partitioning of variance such that it will be possible to decompose variance in overall

performance ratings into within-supervisor and between-supervisor components. Finally, the HLM methodology will allow hypotheses regarding cross-level effects to be tested (e.g., how supervisor characteristics affect the trooper-level relationship between ratings of contextual and overall performance).

The logic of HLM is that both within- and cross-level relationships are modeled by simultaneously estimating two models, a level-1 model, which models relationships within groups, and a level-2 model, which models relationships between groups (Hofmann, 1997). The level-1 model uses the outcome measure for individual  $i$  in group  $j$  as the dependent variable to estimate intercepts and slopes within each group (i.e., intercepts and slopes can differ across groups). In the level-2 model, relationships can be modeled between group-level variables and the level-1 outcome variable as well as between group-level variables and the level-1 slopes and intercepts. In other words, the level-2 model can be used to answer two questions, “what is the effect of a higher level unit characteristic on a lower level outcome?” and, “what is the influence of a higher level unit characteristic on the relationship between lower level variables?” (Klein & Kozlowski, 2000, p. 230). In the proposed study, employees are grouped within the supervisors to which they report. That is, the level-1 model uses the outcome measure for trooper  $i$  rated by supervisor  $j$  as the dependent variable to estimate intercepts and slopes within each supervisor. Supervisor-level variables are added in the level-2 model. The following discussion outlines the proposed models to be tested, beginning with a fully unconditional model in which no predictors are included at level 1 or level 2,



followed by models which incorporate predictors at level 1 only, and finally models which incorporate predictors at both levels.

### *Fully Unconditional Model*

The simplest possible hierarchical model, the fully unconditional or “null” model, is typically assessed as the first step in HLM analysis (Raudenbush & Bryk, 2002). The model is equivalent to a one-way ANOVA with random effects and can be used to calculate a point estimate for the grand mean of the criterion, as well as parameters representing the variability of the criterion variable at both levels. The model is said to be fully unconditional in the sense that no predictor variables are entered into the equations at either level. Instead, the level-1 model predicts the outcome for employee  $i$  using only the intercept,  $\beta_{0j}$ , where the  $\beta_0$  is the notation used for intercepts and the subscript  $j$  refers to the supervisor to which the employee reports. The intercept which is the mean outcome for the  $j$ th supervisor (i.e.,  $\mu_{Y_j}$ ), and is accompanied by a level-1 residual term,  $r_{ij}$  for employee  $i$  rated by supervisor  $j$ , which is assumed to be normally distributed with a mean of zero and a variance,  $\sigma^2$  that is constant across level-1 units (employees). The level-1 formula is given as

$$Y_{ij} = \beta_{0j} + r_{ij}. \quad (1)$$

where  $Y_{ij}$  is the level-1 outcome measure (overall performance rating) for employee  $i$  rated by supervisor  $j$ . The accompanying level-2 model is as follows,

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad (2)$$

where  $\gamma_{00}$  is the grand mean of the outcome measure (due to the grand mean centering procedure described previously) and  $u_{0j}$  is the random effect associated with supervisor  $j$  which is assumed to have a mean of zero and variance equal to  $\tau_{00}$ . Thus, the fully unconditional model yields  $\gamma_{00}$ , a point estimate for the grand mean of the criterion,  $\sigma^2$ , the within-supervisor (level-1) variability, and  $\tau_{00}$ , the between-supervisor (level-2) variability. In the language of fixed and random effects, the model contains one fixed effect,  $\gamma_{00}$ , the overall intercept, and two random effects,  $\tau_{00}$  and  $\sigma^2$ , the variability between and within supervisors, respectively.

For the present study, where ratings of individual troopers are grouped by the supervising sergeant providing the rating, the proposed unconditional model is:

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + r_{ij} \quad (3)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}. \quad (4)$$

In this model, the dependent variable is the rating of individual troopers' overall performance. At level 1, this rating is expressed as the sum of an intercept for supervisor  $j$  ( $\beta_{0j}$ ) and a random effect ( $r_{ij}$ ) associated with the  $i^{\text{th}}$  trooper rated by the  $j^{\text{th}}$  supervisor. At level 2, the intercept is expressed as the sum of the grand-mean rating ( $\gamma_{00}$ ) and a random effect ( $u_{0j}$ ) associated with the  $j^{\text{th}}$  supervisor.

An initial test of the unconditional model indicated that there is significant between-supervisor variance component ( $\tau_{00} = 0.10$ ,  $z = 3.06$ ,  $p < .01$ ), as well as a significant within-supervisor component ( $\sigma^2 = 0.39$ ,  $z = 11.24$ ,  $p < .01$ ) with regard to the overall performance ratings of individual troopers. That the variance is statistically significant at both levels

indicates that the development of models that include individual- and group-level predictors of overall performance is justified. Before turning to those models, however, it is common practice to use these variance components to estimate the intraclass correlation coefficient.

#### *Intraclass Correlation Coefficient (ICC)*

A useful coefficient associated with the unconditional model is the intraclass correlation coefficient ( $\rho$ ), calculated as:

$$\rho = \tau_{00} / (\tau_{00} + \sigma^2). \quad (5)$$

The coefficient provides a measure of the between-supervisor variance (i.e., variance between level-2 units) as a proportion of the total variability. Once the ICC is calculated, the percentage of variance that is attributable to within-supervisor variance may be calculated by simply subtracting the coefficient from one, (i.e.,  $1 - \rho$ ). In the current study, dividing the between-supervisor variance by the total variance (Equation 5) yields an ICC equal to .202, which is to say that 20.2% of variability in ratings of overall performance occurs between supervisors, while 79.8% is within supervisors.

#### *Proposed Level-1 Models*

Raudenbush and Bryk (2002) recommend that model building begin with the level-1 model. In this stage, the researcher tests models which are unconditional at level 2 (i.e., no level 2 predictors are included) in order to identify the most appropriate level-1 model. As such, this section describes models that build on the fully unconditional model by adding predictors at level 1 only.

*Random-coefficients regression model.* Having confirmed that there is variability in the level-1 performance ratings, it is appropriate to add predictors of these ratings to the baseline model, the first model to be tested is a random-coefficients regression model. While Hypothesis 1 specifies main effects for two level-1 predictors, task performance contextual performance, the model will be tested using the recommended “step-up” approach (Raudenbush & Bryk, 2002), in which the analyst first tests a single-predictor model and then proceeds by adding variables one at a time. If a variable does not have a significant fixed effect and there is no evidence of slope heterogeneity, it can be omitted from the model.

The proposed single-predictor model, in which task performance (TP) is entered as the sole predictor of overall performance, can be written as follows:

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + r_{ij} \quad (6)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (7)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (8)$$

Comparing Equation 6 to Equation 3 (level 1 of the fully unconditional model), the dependent variable (overall performance rating), the intercept ( $\beta_{0j}$ ), and the random effect for troopers within supervisors ( $r_{ij}$ ) are all carried over. With the addition of the predictor, task performance,  $\beta_{0j}$  is now interpreted as the expected overall performance for an individual with an average task performance rating. Added to the equation is the slope for task,  $\beta_{1j}$ , which becomes a dependent variable at level 2 (Equation 8). Note that  $\gamma_{00}$  is now interpreted as the average supervisor mean performance rating, controlling for individual task ratings. A significant  $t$ -statistic associated with  $\gamma_{10}$  would indicate that the estimated average slope

relating trooper task performance and overall performance is significantly different from 0, thereby providing partial support for Hypothesis 1.

Because this model allows both the intercept and the slope to vary randomly, a new random effect appears in Equation 8,  $u_{1j}$ , which represents the unique increment to the slope associated with supervisor  $j$  and has variance equal to  $\tau_{11}$  (Raudenbush & Bryk, 2002). If, as predicted by Hypothesis 2(a), the slope for task performance does not vary across supervisors, the results of this model will indicate that  $\tau_{11}$  (which is the term representing the variance of  $u_{1j}$ ) is not significant. In this case, it may be that a simpler model in which slopes do not vary across supervisors (i.e., the effect  $u_{1j}$  is omitted) may provide a better fit of the data.<sup>4</sup> This hypothesis can be tested by comparing the Akaike's Information Criteria fit indices associated with the two models.

After selecting the appropriate task performance model, the hypothesis that contextual performance (CP) is also significantly related to overall performance will be tested. So as to test the hypothesis as worded, and not the hypothesis that contextual performance accounts for variance in overall performance above and beyond that accounted for by task performance, task performance will be removed from the equation. Hence, the new model will be written as:

$$\text{Level 1: } Perf_{ij} = \beta_{0j} + \beta_{1j}(CP) + r_{ij} \quad (9)$$

---

<sup>4</sup> Note that by constraining the TP slope across supervisors, the model is technically no longer a random coefficients regression model, but actually a one-way ANCOVA with random effects.

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (10)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}. \quad (11)$$

Full support for Hypothesis 1, which states that both task performance and contextual performance ratings will be significantly related to ratings of overall performance, would be provided by significant  $t$ -statistics associated with coefficients  $\gamma_{10}$  and  $\gamma_{20}$ . Research Question 1, which concerns the variance in overall performance ratings attributable to task and contextual performance, can also be tested at this point. For example, to assess the amount of employee-level variability in overall performance that task performance accounts for, the calculation would be as follows:

$$R^2_{\text{within}} = \frac{\sigma_{uc}^2 - \sigma_c^2}{\sigma_{uc}^2}, \quad (12)$$

where  $\sigma_{uc}^2$  is the residual level-1 variance in the unconditional model (i.e., the model with no predictors) and  $\sigma_c^2$  is the residual level-1 variance for the model in which task performance is entered as a predictor at level 1. The result is the proportion of level-1 variability in overall performance that can be attributed to task performance ratings.

Hypothesis 2(b) states that the slope for contextual performance will vary across supervisors. This variation is allowed by the inclusion of the random effect  $u_{2j}$  in Equation 11. A significant  $z$ -statistic for  $\tau_{11}$  (i.e., the variance of  $u_{1j}$ ) would indicate support for this hypothesis and also the appropriateness of testing models in which level-2 variables are added to explain this variance (i.e., cross-level moderators).

By comparing the  $\sigma^2$  estimate for this model with the  $\sigma^2$  estimate for the fully unconditional model, it will be possible to determine the proportion of the within-supervisor variance accounted for by the two predictors together. Presuming that significant within-supervisor variance in overall performance remains unexplained after the addition of task and contextual performance,  $q$  additional level-1 predictors will be added to the model, following the same “step-up” procedure. As before, if the  $\tau_{qq}$  associated with a slope is not significant, an alternative model where the slope is not allowed to vary will be tested. If a  $\gamma_{q0}$  is non-significant, that predictor will be removed from the model altogether.

Hypotheses 3 and 4 state that trooper tenure will be positively related to overall performance and that tenure will moderate the relationship between contextual performance and overall performance. The model to test these hypotheses is:

$$\text{Level 1:} \tag{14}$$

$$Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + \beta_{3j}(tenure) + \beta_{4j}(CP * tenure) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \tag{15}$$

$$\beta_{1j} = \gamma_{10} \tag{16}$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \tag{17}$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \tag{18}$$

$$\beta_{4j} = \gamma_{40} + u_{4j}. \tag{19}$$

Support for Hypothesis 3 would be provided by a significant  $t$ -statistic for  $\gamma_{30}$  while a significant  $t$  for  $\gamma_{40}$  would provide support for Hypothesis 4. An identical procedure will be used to investigate Research Questions 2 regarding opportunity to observe.

#### *Proposed Level-2 Models*

Should the results of the random coefficients regression models indicate significant variance in the slope and/or intercept terms across supervisors, models which incorporate level-2 predictors of this variance will be tested. For simplicity, the following models will include the level-1 effects as they are outlined in Hypotheses 1-4. The actual model specification will, of course, be informed by the results of the level-1 investigation.

*Intercepts-as-outcomes model.* Research Question 3 specifies a cross-level main effect, namely, that a supervisor's amount of experience (operationalized as self-reported time in current rank) will be positively related to ratings of overall performance. Whereas the models up to this point have been unconditional at level 2 in that no attempt was made to explain variance in slopes or intercepts, here the level-2 predictor, experience, is added to the intercept equation (Equation 21):

Level 1: (20)

$$Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + \beta_{3j}(tenure) + \beta_{4j}(CP * tenure) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(experience) + u_{0j} \quad (21)$$

$$\beta_{1j} = \gamma_{10} \quad (22)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (23)$$



$$\beta_{3j} = \gamma_{30} \quad (24)$$

$$\beta_{4j} = \gamma_{40}. \quad (25)$$

A significant  $t$ -statistic for  $\gamma_{01}$  would indicate that supervisor experience does have an effect on overall performance ratings.

*Slopes-as-outcomes model.* Presuming that significant between-supervisor variance remains unexplained, the cross-level moderation effects specified by Research Question 4 and Hypotheses 5 and 6 will be tested. For Research Question 4, which asks whether the relationship between contextual performance and overall performance depends on supervisor experience, the intercepts-as-outcomes model is modified to include experience in the contextual performance slope equation (Equation 29) as well as the intercept equation (Equation 27):

$$\text{Level 1:} \quad (26)$$

$$Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + \beta_{3j}(tenure) + \beta_{4j}(CP * tenure) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(experience) + u_{0j} \quad (27)$$

$$\beta_{1j} = \gamma_{10} \quad (28)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(experience) + u_{2j} \quad (29)$$

$$\beta_{3j} = \gamma_{30} \quad (30)$$

$$\beta_{4j} = \gamma_{40}. \quad (31)$$

A significant  $t$ -statistic associated with  $\gamma_{21}$  would indicate that experience does moderate the relationship between ratings of contextual and ratings of overall performance.

The final hypotheses assert that characteristics of the district a particular supervisor is responsible for (namely, the district's number of interstates and total population) will moderate the contextual-overall performance relationship. While these hypotheses will be evaluated in the same "step-up" fashion described throughout this proposal, the full model, which includes both of these effects along with those specified in Hypotheses 1-4, is written as follows:

$$\text{Level 1:} \quad (32)$$

$$Perf_{ij} = \beta_{0j} + \beta_{1j}(TP) + \beta_{2j}(CP) + \beta_{3j}(tenure) + \beta_{4j}(CP * tenure) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(experience) + \gamma_{02}(interstates) + \gamma_{02}(pop) + u_{0j} \quad (33)$$

$$\beta_{1j} = \gamma_{10} \quad (34)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{22}(interstates) + \gamma_{23}(pop) + u_{2j} \quad (35)$$

$$\beta_{3j} = \gamma_{30} \quad (36)$$

$$\beta_{4j} = \gamma_{40}. \quad (37)$$

## Appendix D

### Dissertation Proposal Document Tables

*Table D.1*

Self-Reported Time in Rank for Sergeant Sample

	Frequency	Percent of Total
0-3 months	5	5.56%
4-6 months	3	3.33%
7-9 months	4	4.44%
10-12 months	2	2.22%
1 year	14	15.56%
2 years	12	13.33%
>3 years	50	55.56%

(Appendix D continues)

Appendix D (cont.)

*Table D.2*

Demographic Characteristics of Trooper Sample

	Frequency	Percent of Total
Race		
African American	30	8.96%
American Indian	13	3.88%
Asian American	2	0.60%
Caucasian	289	86.27%
Hispanic	1	0.30%
Gender		
Female	6	1.79%
Male	329	98.21%

(Appendix D continues)

# Appendix D (cont.)

*Table D.3*

EFA Results for Contextual Performance Items (Promax-Rotated Factor Pattern)

	Factor1	Factor2	Factor3
CY4	<b>74</b>	22	-12
A2	<b>73</b>	12	6
A4	<b>68</b>	1	18
CY3	<b>63</b>	14	13
CY5	<b>61</b>	13	18
A1	<b>59</b>	11	8
A3	<b>58</b>	0	31
CY1	<b>58</b>	31	-7
CY2	<b>57</b>	17	14
A5	<b>46</b>	-4	41
S2	8	<b>86</b>	-5
S3	5	<b>83</b>	0
S5	14	<b>74</b>	0
S4	21	<b>65</b>	0
S1	11	<b>58</b>	-1
CV4	6	-13	<b>68</b>
CV2	25	-16	<b>65</b>
CV3	25	-12	<b>61</b>
CS3	-17	31	<b>57</b>
CS5	5	40	<b>56</b>
CS4	1	31	<b>54</b>
CV1	9	0	<b>54</b>
CS1	-5	36	<b>48</b>
CS2	1	15	<b>32</b>

(Appendix D continues)

Appendix D (cont.)

*Table D.4*

Length of Supervisory Relationships for the Sample

	Frequency	Percent of Total
0-3 months	37	11.04%
4-6 months	25	7.46%
7-9 months	21	6.27%
10-12 months	14	4.18%
1 year	70	20.90%
2 years	87	25.97%
>3 years	81	24.18%

(Appendix D continues)

## Appendix D (cont.)

*Table D.5*

Sergeants' Reported Opportunity to Observe Trooper Performance

	Frequency	Percent of Total
Daily	36	10.75%
Weekly	222	66.27%
Monthly	65	19.40%
Every 3 months	8	2.39%
Every 6 months	3	0.90%
Once every year	1	0.30%
Once every few years	0	0.00%

*Table 1*

Fit Indices for Task Performance Models

	Random intercepts and slopes	Random intercepts only
-2 Res Log Likelihood	318.1	319.7
AIC*	326.1	323.7
AICC*	326.2	323.7
BIC*	335.9	328.6

Note: \* indicates that smaller values are preferred.



Table 2

Results of Task and Contextual Performance Model

Fixed Effect	Coefficient	SE	df	<i>t</i> Value	<i>p</i>
Intercept, $\gamma_{00}$	5.002	0.026	85	195.26	<.0001
Task performance, $\gamma_{10}$	0.891	0.043	236	20.49	<.0001
Contextual performance, $\gamma_{20}$	0.156	0.041	236	3.81	.0002
Random Effect	Coefficient	SE		<i>Z</i> Value	<i>p</i>
$\tau_{00}$	0.019	0.010		1.94	0.0259
$\tau_{10}$	0.004	0.008		0.44	0.6592
$\tau_{11}$	0.035	0.014		2.58	0.005
$\sigma^2$	0.107	0.011		9.79	<.0001

*Notes.*  $\tau_{00}$  = unconditional variance in the level-1 intercepts;  $\tau_{10}$  = unconditional covariance

between the level-1 slopes and intercepts;  $\tau_{11}$  = unconditional variance in the level-1 slopes.

$\sigma^2$  = residual level-1 variance.