

MULTIVARIATE DYNAMIC CRITERIA:  
A PROCESS MODEL OF JOB PERFORMANCE

BY

SETH MICHAEL SPAIN

DISSERTATION

Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Psychology  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Professor Fritz Drasgow, Chair, Director of Research  
Professor Charles L. Hulin, *emeritus*  
Professor Brent W. Roberts  
Assistant Professor Daniel A. Newman  
Assistant Professor Terry L. von Thaden

# ABSTRACT

As predictive validity is so generally established using bivariate correlations or multiple regressions job performance must usually be reduced to a single variable. This is in spite of considerable evidence that job performance is a multidimensional construct composed of several facets. Additionally, the predictive validity of ability tests declines with increasing time from testing, suggesting that job performance may best be viewed as a developmental construct. When both of these premises are true, it is possible that different developmental trends in job performance facets are obscured by considering only univariate performance. The purpose of this paper is to explore how to predict the development of different dimensions of job performance over time. We refer to this situation as validation for multivariate dynamic criteria. Viewing predictive validity in this way necessitates a process model for the domain of job performance. This paper develops a theoretical process model that draws together concepts from socialization and the literature on skill acquisition (Ackerman, 1987; Murphy, 1989) and presents a multivariate multilevel model for multiple performance facets measured at 4 times. The model was fitted to data from a 12-week long police training academy using the SAS and R systems. Cognitive ability and Big 5 personality traits are used to predict slopes and intercepts for the performance criteria. Changes in the criteria are mapped onto changes in identity and motivation mechanisms. The pattern of findings is generally consistent with responsive development and socialization effects.

*For my grandparents, Donald and Mary Loos.*

# ACKNOWLEDGEMENTS

This project would not have been possible without the advice, care, and support of my adviser, Fritz Drasgow, and my committee members, Dan Newman, Brent Roberts, and Terry von Thaden, and Chuck Hulin. I must add extra thanks to Chuck for being such a committed teacher and mentor, in spite of his retirement. I also have to thank many of my peers who helped me to develop the ideas expressed here, especially Reeshad Dalal, Andy Miner, Michael Bashshur, Serena Wee, and Chris Nye. I'd also like to extend additional thanks to Louis Tay and Serena Wee, again, for assistance in learning *R*. A special thank you is offered to Larry Hubert for introducing me to the  $\text{\TeX}$  typesetting system and multivariate methods and to Carolyn Anderson, for her introduction to multilevel models. I also want to thank my mother and father, Vida and Steven Spain, for all of their kindness, support, and love.

I would like to thank the Director and staff of the Police Training Institute at the University of Illinois for their assistance with this project, particularly Krystal Fitzpatrick, Christine Somers, Chuck Deakin, Melissa Zindars, Evelyn Wilhelm, and Mike Miller.

I also need to thank the small army of undergraduate research assistants who made it possible to complete this project: Daniel Kougias, Melissa Jones, Rajiv Khattar, Youbin Cha, Yusra Al-Shawaf, Dahye Kim, Jun Meng, Gyu San Lee, Manwen Guo, and Norah Linthicum. I would especially like to thank Jill Fleming for her assistance with this and several other projects.

Finally, I would like to thank Emily Love, for all of her support and love.

# CONTENTS

LIST OF TABLES . . . . .	vii
LIST OF FIGURES . . . . .	viii
Chapter 1 INTRODUCTION . . . . .	1
Chapter 2 LITERATURE REVIEW . . . . .	6
2.1 A behavioral definition of job performance . . . . .	6
2.2 Operationalizing performance criteria . . . . .	8
2.3 Dimensional problems of criteria . . . . .	9
2.4 Implications of the multidimensional nature of job performance	11
2.5 Implications of dynamic criteria . . . . .	13
2.5.1 Multilevel perspectives on performance over time . . . .	15
2.6 Validation: A causal model of job performance . . . . .	17
2.7 Validation evidence: A review . . . . .	18
2.7.1 Cognitive Ability . . . . .	19
2.7.2 Personality . . . . .	20
2.8 Lifespan development perspectives on the performance repertoire	22
2.8.1 Identity and self-regulation: Mechanisms underlying	
behavior in performance episodes . . . . .	23
2.8.2 Building a framework for the validation of multi-	
variate dynamic criteria . . . . .	27
2.9 The current study . . . . .	28
2.9.1 Research questions . . . . .	28
2.9.2 The causal model . . . . .	29
2.9.3 Hypotheses . . . . .	30
2.9.4 Ancillary predictions . . . . .	33
Chapter 3 STUDY OVERVIEW . . . . .	34
3.1 Sample . . . . .	34
3.2 Procedure . . . . .	35
3.3 Measures . . . . .	35
Chapter 4 ANALYTICAL STRATEGY . . . . .	37
4.1 Static hypotheses . . . . .	37
4.2 Dynamic hypotheses . . . . .	37

Chapter 5	RESULTS . . . . .	41
5.1	Descriptive Statistics . . . . .	41
5.2	Hypothesis testing . . . . .	44
5.3	Slopes as predictors . . . . .	61
5.4	Self-Instructor Correlations. . . . .	62
Chapter 6	DISCUSSION . . . . .	69
6.1	Limitations . . . . .	72
6.2	Future Directions . . . . .	73
6.3	Conclusions . . . . .	74
REFERENCES	. . . . .	80
VITA	. . . . .	90

# LIST OF TABLES

5.1	Descriptive Statistics for level 1 variables for Weeks 3, 6, 9 and 12. . . . .	42
5.2	Descriptive Statistics for between person variables. . . . .	42
5.3	Correlations for major level-2 predictors for weeks 3 and 6 . .	43
5.4	Correlations with performance for major level-2 predictors for weeks 9 and 12 . . . . .	44
5.5	Confirmatory factor analyses . . . . .	45
5.6	Univariate change over time in performance indicators . . . .	51
5.7	ICCs, Likelihood Ratio tests, and fixed effect estimates of performance measures . . . . .	52
5.8	ICCs, Likelihood Ratio tests, and fixed effect estimates of time-varying predictor change . . . . .	54
5.9	REML estimates of fixed effects for unique effects of time . . .	56
5.10	REML estimates of random effects for individual differences in growth curves . . . . .	63
5.11	Likelihood ratio tests for unique effects of within-person covariates. . . . .	64
5.12	REML estimates of fixed effects for within-person covariates .	65
5.13	Likelihood ratio tests for hypothesis between-person predictors	66
5.14	REML estimates of fixed effects for between-person predictors	67
5.15	Correlations for intercepts and slopes predicting later outcomes	67
5.16	Correlations between self- and instructor ratings of performance	68

# LIST OF FIGURES

2.1	Data Relations Box . . . . .	10
2.2	Model Schematic . . . . .	24
2.3	Basic causal model of performance ratings . . . . .	26
2.4	Example individual growth model of performance ratings . . . . .	30
5.1	Item cluster analysis for Week 3 . . . . .	47
5.2	Item cluster analysis for Week 9 . . . . .	48
5.3	Item cluster analysis for all times . . . . .	49



# Chapter 1

## INTRODUCTION

Personnel selection and the measurement of job performance have long occupied the scholarly attention of industrial and organizational psychologists. The development of selection measures and procedures has been a major focus of research and applied work for the last century (Schmidt & Hunter, 1998). The advent of meta-analytic procedures has helped clarify issues regarding how well measures of individual differences can predict job performance, leading to the consensus that cognitive ability tests are the best predictors of overall job performance (Ree & Earles, 1994; Schmidt & Hunter, 1998). In addition, personality measures, such as tests of conscientiousness can also provide reasonable prediction of job performance (Barrick & Mount, 1991). These personality measures can add substantial incremental validity over cognitive tests (Schmidt & Hunter, 1998).

Much of the emphasis in studying the performance-prediction problem has been on developing and refining predictors, with too little attention paid to the performance criteria these measures are supposed to predict (Campbell, 1990). Reviews of the job performance literature often call for greater study of the criteria themselves (Borman, 1991; Borman, Hanson, & Hedge, 1997; Ghiselli, 1956; Weitz, 1960), giving particular attention to the latent dimensionality and temporal aspects of the job performance construct.

This so-called *criterion problem* has long plagued organizational scientists (Austin & Villanova, 1992). The criterion problem arises because there is considerable difficulty in defining the boundaries of the job performance construct and then translating that definition into measurement instruments. This paper begins by examining job performance in detail, constructing a multidimensional and developmental model of job performance. Job performance is defined as a broad behavioral repertoire (Humphreys, 1992; Motowidlo, Borman, & Schmit, 1997). It is expected that the behaviors in the repertoire will be reflected in measurements of job performance crite-

ria, which should be sampled broadly (e.g., Hanisch, Hulin, & Roznowski, 1998). Evidence that measurements of performance criteria demonstrate a multidimensional structure is then examined. Next we discuss how treating multidimensional performance facets as unidimensional, as is often the case, can cause severe problems in estimating validity (Murphy & Shiarella, 1997; Murphy, 2009). Furthermore, the criterion is not necessarily static, but may be distributed within individuals over time (Austin, Humphreys, & Hulin, 1989; Henry & Hulin, 1987; Hulin, Henry, & Noon, 1991; Humphreys, 1960; Kane, 1986).

This vantage point provides a view of job performance as an outcome of the general adult development process. Therefore, it is expected that both the size and contents of the repertoire should change over time as individuals selectively optimize their efforts towards work role-consistent goals (e.g., Baltes, 1997; Day, Harrison, & Halpin, 2009). The evidence that performance demonstrates such a developmental trend is examined, with particular emphasis given to the problems such changes pose for making valid inferences from predictors measured at a single point in time (Hofmann, Jacobs, & Baratta, 1993; Hulin, Henry, & Noon, 1990).

Given the above issues of multidimensionality and temporal instability, current models of predictive validity for job performance criteria are inadequate; however, the validation of selection measures is simply an application of the scientific method (Cronbach & Meehl, 1955; Messick, 1989; 1995). A claim that a test is valid means that inferences made about scores on the test are accurate for a particular purpose or set of purposes (Angoff, 1988). Test validation involves determining that the instrument used measures the construct it is intended to measure, and that it demonstrates theoretically appropriate external relations, specifically, predicting criterion measurements. Therefore, a claim of predictive validity is an implied causal claim; predictive validity is a directed association between some predictor construct and some outcome construct (Binning & Barrett, 1989).

The approach advocated in this thesis builds on previous causal models of job performance (Hunter, 1983; Schmidt, Hunter, & Outerbridge, 1986). Such models explicitly treat the performance prediction problem as one of causal inference. Causal models of job performance attempt to extend validation research by capturing the mediating processes that transmit the causal influence of the personal attributes measured by the selection system to per-

formance behaviors. As a causal hypothesis, validation designs should maximize the strength of this inference. Hence, designing a validation study for a measure requires an adequately broad sampling of performance indicators, and appropriate timing of predictor and criterion measurements. Such a study should also specify appropriate mediating variables, so that a causal model of the multivariate dynamic processes underlying the performance repertoire may be tested; this design allows relatively strong inferences about the process to be made (e.g., Heckman, 2008; Pearl, 2009; cf. Smith, 1976). Validation designs are usually incapable of controlling for the variance of all possible third variables, so the “causal” models discussed throughout do not truly demonstrate causation. They merely show whether the data are consistent with the specified theoretical model or not. Many other causal structures are possible.

The current study addresses the process of establishing validity in the context of these multidimensional and changing job performance criteria. Validity will be demonstrated by examining a process model for the multidimensional performance repertoire. This model is built by drawing on the adult development literature (e.g., Baltes, 1987; 1997; Day et al., 2009). Taking this perspective as a guide, individuals are viewed as being fundamentally multilevel phenomena themselves. The iceberg metaphor will be used for individuals: the behavioral repertoire is merely an observable, surface level manifestation. Beneath that surface exists a variety of affective and regulatory mechanisms, for example, goal orientation and self-efficacy, which drive behavior at a given time. At the deepest level, far below the surface, are identity mechanisms, such as the self-concept or the individual’s view of his or her own personality traits and motives, which are the core developmental constructs considered in the proposed model. This hierarchy of mechanisms is consistent with the *Neo-socioanalytic model* of personality (Roberts, 2006). These perspectives are drawn together by invoking concepts of socialization and corresponsive development (Roberts, 2006; Roberts & Caspi, 2003). Individuals will invest their limited resources to optimize performance in specific roles, and it is expected that they will choose investment patterns most in line with their dispositional preferences (c.f., Baltes; 1997)

Each of these levels is dynamic, and change in one level should feed through to the other levels, but the levels need not change at the same rate. It is expected that these levels will interact in spiraling patterns (Day

et al., 2009; Day & Sin, 2009). For example, as an individual attempts a new work task and succeeds, that success will reinforce the individual's perception of him- or herself as a competent worker, which should in turn foster positive beliefs about work skills, which should lead to further attempts at new tasks. On the other hand, if the worker attempts new tasks and fails them, the weight of this failure or punishment from the environment (e.g., supervisor reprimands the worker) should foster negative beliefs about the worker's skills and negatively impact his or her self-views, potentially leading to a downward performance spiral. Such environmental feedback could also encourage certain workers to blame their tools or develop the idea that the task is meaningless or impossible. This study will focus on internal attribution processes, but there are other belief mechanisms that could lead to similar outcomes.

Furthermore, if the performance repertoire is indeed multidimensional, the influence of changes in the deeper levels of personality organization may even be in opposite directions for different aspects of performance. For instance, broad trait measurements may obscure trends in their lower-order facets. Two major facets of the personality trait Extraversion are Social Dominance and Social Vitality. Mean levels of Social Dominance appear to increase throughout the lifespan, whereas Social Vitality seems to peak in early adulthood and then decline (Roberts, Walton, & Veichtbauer, 2005). So, change in a deep level construct may filter up to the surface with differing manifestations in different domains of behavior. For instance, someone may become more agreeable as she ages, which may allow for more fluid social interactions with peers, but may make it more difficult to provide critical feedback to subordinates.

From a perspective based on adult development, the validation of selection tools requires appropriate timing of criterion measurements, in order to capture the variability and trajectory of performance facets. For instance, estimating a true intercept for an individual growth model would require a "time zero" measurement, a time when the participant has effectively no experience with the job and no appreciable job knowledge or skill. Repeated measurements of the criteria must then be taken with a frequency and patterning that are appropriate to model the trajectory of job performance. In order to correctly time these measurements requires a relatively strong underlying theory regarding the developmental process of job performance

behaviors.

In summary, this paper presents a viewpoint of job performance as a repertoire of behavior, which is a surface level outcropping of a particularly important role-identity, that of the worker. This system of behaviors is open and dynamic, and the development of the repertoire rests on changes in underlying identity and self-regulatory mechanisms. This system will be explored by examining the trace evidence of behavior that is captured in multivariate dynamic criterion measurements (e.g., Dalal & Hulin, 2008), in that several facets of the performance repertoire are measured at several points in time. The development of the repertoire will be mapped onto changes in identity and regulatory mechanisms with self-regulation operationalized by goal orientation, self-efficacy, affect, and motives for doing the job, and identity operationalized by personality trait ratings. Furthermore, these changes will be related to individual differences that can be measured in a selection setting, specifically cognitive ability and personality traits. The validity of those individual difference measures is expressed by evaluating a process model of the development of the job performance repertoire.

## Chapter 2

# LITERATURE REVIEW

### 2.1 A behavioral definition of job performance

Let us consider a definition of job performance. Thorndike's (1949) ultimate criterion provides a historically reasonable starting point (e.g., Borman, 1991): the sum of all contributions that an individual makes to the workplace. The example offered by Thorndike is that of an insurance salesperson who sells all the policies it is possible to sell, never allows those policies to lapse, and continues to do so for very many years. The ultimate criterion is ultimate in the sense that it is the absolute goal of any selection or training program. It is far abstracted from the particular performance measurements that practicing managers will employ to rate their subordinates. Yet such abstraction is useful for organizational scientists to derive a meaningful operationalization of the job performance construct. Viewing performance from the standpoint of the ultimate criterion shows that we should be concerned that our criterion measures show sufficient breadth, covering the major domains of contribution an employee can make, and that the measures chosen reflect the lifetime contribution an employee makes.

We will think of the job performance repertoire analogously to Humphrey's (1992) definition of intelligence: the set of learned behaviors comprised of knowledge, skills, abilities, and other characteristics considered to be intellectual. This paper will consider job performance as a behavioral construct (e.g., Motowidlo et al., 1997): job performance is a repertoire of behaviors that are needed to execute the tasks, duties, responsibilities, and other obligations that are relevant to achieving organizational goals (Astin, 1964; Murphy, 1989). This set is fuzzy in that such behaviors need not belong exclusively to the domain of job performance; they may be necessary to perform in other roles, such as playing sports, raising children, or enjoying leisure ac-

tivities. Which behaviors constitute work for some individuals may be play for others.

Naturally, all behaviors have causes, and job performance behaviors are no different (e.g. Wernimont & Campbell, 1968). Knowledge and skill are important determinants of job performance behaviors (Campbell, 1990; Motowidlo et al., 1997). Furthermore, motivational or habitual variables also determine performance behaviors. Several studies of the antecedents of performance support this contention (Borman, White, Pulakos, & Oppler, 1991; Hunter, 1983; Schmidt et al., 1986). These studies show that the influence of traits such as cognitive abilities or personality characteristics on job performance ratings is mediated by the job knowledge and skills that the individual has developed. Motowidlo and colleagues described these mediating constructs as the outcome of developmental transactions between the individual's deep characteristics and environmental pressures; they are "characteristic adaptations", in the terminology of McCrae and Costa, 1996.

Behaviors are not necessarily fungible. For example, as a professor, when one is engaged in a teaching episode such as lecturing, it is not possible to simultaneously be conducting research. Hence, an individual does not engage in all of these behaviors at all times. Instead, job performance is composed of many discrete behavioral episodes (Beal, Weiss, Barros, & MacDermid, 2005; Motowidlo et al., 1997). These episodes can be scaled in terms of their dollar value to the organization (Motowidlo et al., 1997). But, we must note that not all behavioral episodes enacted at work will be relevant to organizational goals, some may be neutral. Also, some behaviors will provide positive contributions to the organization's technical core (e.g., producing a document, selling a car), while other behaviors may be positive to the social environment that supports that technical core (e.g., assisting a colleague with a difficult analysis, mentoring a junior salesperson). Other behaviors may be negative or destructive (e.g., conducting unethical research, joyriding in a dealer vehicle), and some may signal poor working conditions or dissatisfaction (e.g., surfing the internet; being late to work). It follows that, while the performance repertoire can be seen as a unidimensional construct at the level of organizational utility, the behaviors themselves reflect potentially many facets.

The behaviors that an individual enacts in any specific behavioral episode are likely to be the product of any number of deep traits and situational con-

straints. Some of these factors are likely to influence chronically accessible behaviors. For example, undoubtedly, the knowledge and skill possessed by the individual is implicated (Campbell, 1990; Motowidlo et al., 1997; Schmidt et al., 1986). Furthermore, motivational variables are necessary (Campbell, 1990). All the skill in the world will not result in performance behaviors if the individual does not choose to apply that skill to the tasks at hand. Work habits may also be essential for describing an individual’s typical levels of performance (Motowidlo et al. 1997). Other factors may be transitory. For instance, the Broaden and Build theory implicates positive emotion in making exploratory behaviors temporarily more accessible (Frederickson, 2001; Seo, Feldman-Barrett, & Bartunek, 2003). Furthermore, the episodic process model suggests that the attentional and regulatory resources available in any given performance episode, resources that are necessary for engaging in work behaviors, will be depleted by off-task attentional demands and the self-regulation activities required by previous episodes (Beal et al., 2005).

## 2.2 Operationalizing performance criteria

Perhaps reflecting a devotion to the ultimate criterion, for many organizational purposes job performance is distilled to a single “criterion”. This may be some single, unidimensional performance indicator, a factor score for some set of performance indicators, or a composite of indicators. Dunnette (1963) disparaged such “distilled essence” criteria, which he viewed as inappropriately combined indicators of performance (i.e., scaled by loadings on the first principal component, regardless of component structure, etc). Dunnette argued that such criteria demonstrate limited construct validity and he failed to see their usefulness in scientific research.

Such “distilled” criteria need not be meaningless. Hulin (1982) presented a higher-order factor model of performance ratings, with a meaningful general factor at the second level. Hulin argued that since ratings reflecting multiple performance dimensions must often be collapsed to make personnel decisions, use of a general factor involves fewer arbitrary decision rules and preserves more information than other methods. Further evidence for a higher-order factor was provided by Viswesvaran, Schmidt, and Ones (2005), who demonstrated the presence of a general factor in performance ratings



even after controlling for halo error. Additionally, Harrison, Newman, and Roth (2006), based on the compatibility principle (Fishbein & Ajzen, 1975, pp. 292 - 298; 352 -353), argued for the use of broad criteria, aggregating many aspects of performance, when using broad attitudes to predict behavior. Note, however, that all of these approaches involve higher-order models; there exist separable dimensions of performance at lower levels. For example, Harrison, Newman, and Roth (2006) used a definition of general performance that is based on lower-order factors of focal performance, citizenship performance, and withdrawal.

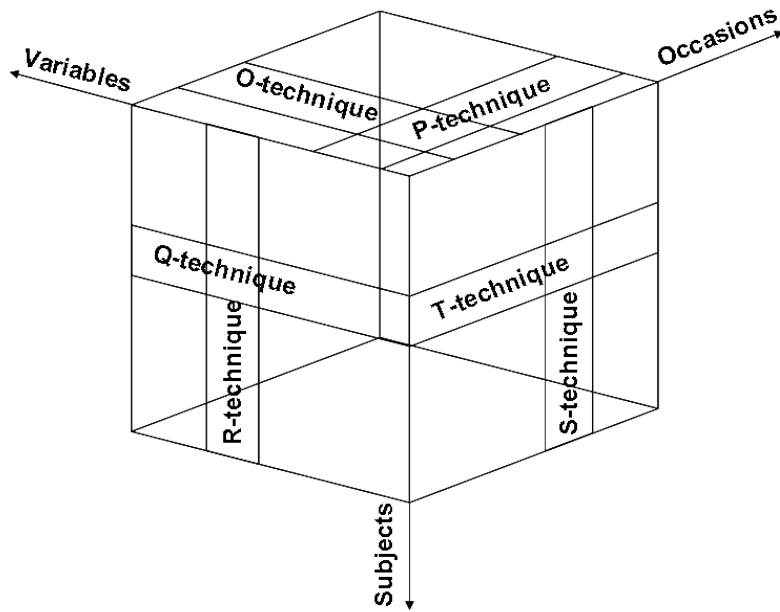
## 2.3 Dimensional problems of criteria

Multifaceted viewpoints have been suggested for operationalizing job performance criteria for many years (e.g., Smith, 1976). For instance, Smith contrasted the ultimate criterion with the measurements that organizations have to make in practice. Practical concerns force decision-makers in organizations to make judgments regarding what is considered important in the measurement of performance (Murphy, 2009; in press). Smith outlines how criteria can be placed into three two-dimensional slabs, ranging from specific to general observations, and immediate to distant time frames. The slabs themselves represent actual performance behaviors, the results of one's behaviors, and the effectiveness of those results.

In parallel, Ghiselli (1956) offered a sophisticated argument that performance criteria demonstrate dimensionality in three ways: (1) statically, such that there are multiple, possibly independent components of performance (for instance, the eight dimensions in the Campbell model); (2) dynamically, with latent temporal trends influencing measurement (for instance linear or quadratic trends in individual growth curves); and (3) individually, such that individuals in the same job may perform in *qualitatively* different ways (for instance, employees have entirely non-overlapping repertoires of job performance behaviors, but both repertoires are effective at executing job functions). The individual dimensions are probably the most difficult dimensions to describe, but Ghiselli provides a good example: two salespersons operate in the same store, one makes many sales generating considerable revenue, while the other generates good will encouraging customers to make pur-

chases throughout the store, but does not make as many direct sales. Both approaches may provide the same economic benefit to the organization, but in *qualitatively* different ways. The data needed to address Ghiselli's conjectures regarding the dimensionality of performance criteria are *three-mode*. Figure 2.1 shows a general outline of factor analytic designs that may be used to address the dimensional problems of job performance criteria (Cattell, 1952; cf., Spain, 2007). In the figure, variables would be indicators of various performance behaviors, for example, objective criteria, ratings of task or contextual performance or absences. Occasions represent different measurement opportunities, for instance, annual or semi-annual performance evaluations.

Figure 2.1: Cattell's Basic Data Relations Box.



Ghiselli's arguments were based on problems identified in the literature

at that time. Empirical study has provided evidence for his claims (Inn, Hulin, & Tucker, 1972). This particular study explored a sample of 184 airline reservation agents measured on 11 performance variables each month for five months. Their analyses recovered a  $4 \times 3 \times 3$  component structure, which can be interpreted as there being 4 “types” of employees who varied in the degree to which they performed 3 different aspects of performance, and that there were three distinct time trends underlying their performance. A study designed and analyzed in this manner provides the most direct test of Ghiselli’s conjecture, but there is further evidence that the performance repertoire is multidimensional and changes over time.

## 2.4 Implications of the multidimensional nature of job performance

While arguments for multiple criteria go back many decades (Smith, 1976), the work of Campbell and colleagues provides some of the best evidence that the job performance repertoire is multidimensional (Campbell, 1990; Campbell, McCloy, Oppler, & Sager, 1993). Using the U.S. Army’s Project A data, these authors uncovered an eight dimensional structure that applies broadly across a sample of jobs. This eight dimensional representation of the job performance repertoire is often referred to as the “Campbell Model”. Specifically, we can discuss two task performance components, job-specific and non-job-specific, which refer to the technical core of the job and technical duties that are dispersed over the organization, respectively. Additional job performance dimensions include written and oral communication proficiency, which are divorced from the quality of the content (which probably falls more under the job-specific task proficiency dimension). We can also measure such dimensions as facilitating team and peer performance, supervision, administration, and maintenance of personal discipline.

Even if one does not use the Campbell model specifically, considerable information about “overall job performance” is lost without considering contextual performance, or the individual’s social and non-technical contributions to the workplace, in addition to task performance (e.g., Borman & Motowidlo, 1997). For example, Motowidlo and Van Scotter (1994) found that ratings of task and contextual performance independently predicted ratings

of overall job performance (see also, Johnson, 2001). Such a finding indicates that organizational decision-makers, in this case, frontline supervisors, care about more than just the core technical contributions an employee makes. In addition to task and contextual performance, we can add refraining from counterproductive behaviors (e.g., Dalal, 2004) or withdrawal behaviors (Hulin, 1991). These dimensions essentially capture whether an employee avoids engaging in behaviors that are harmful to the organization. For instance, Harrison and colleagues (2006) modeled “general”, or overall, performance with focal (task) performance, citizenship (contextual) performance, and withdrawal, as its lower-order indicators.

Ignoring the dimensionality of job performance has consequences for validity inferences (Murphy & Shirella, 1997). These authors employed Monte Carlo simulations of a selection battery consisting of cognitive ability and personality tests in predicting composites of task and contextual performance. They found that the *effective weights* given to the different performance measures accounted for 34% of the variance in predictor battery validity. By effective weights, we mean both the nominal weights assigned by organizational decision-makers (e.g., “We value task performance twice as highly as contextual performance in this organization”) combined with the actual variability of the performance facet. They argued that an organization’s operational definition of performance is a “true and important source of variability in validity” (p. 823). Murphy (2009) would go on to argue that an organization’s values are an inseparable aspect of such decisions. This statement can be seen as a codification of the above; organizational decision-makers encode what they care about, what they value implicitly or explicitly, in their weighting of performance criteria. What aspects of performance are measured and how that information is combined serves to create the incentive structure of the organization, which in turn has natural implications for the behaviors individuals engage in.

Murphy and Shirella also found that an additional 23% of the variance in validity was accounted for by the weights assigned to predictors. Thus, it seems wise to examine the multivariate nature of the predictor-criterion relationships. If this is not properly done, operational validity obtained by considering the bivariate relationships may be much lower than can be obtained for even the same set of predictors and criteria in an appropriately multivariate fashion.

## 2.5 Implications of dynamic criteria

Beyond the potential problems associated with multivariate criteria, there may be additional complications due to criteria that change over time. Early evidence of the dynamic nature of performance criteria was provided by Ghiselli and Haire (1960), who showed that selection test validity fluctuates over time. The term *dynamic criteria* is often used to describe the general decrease in predictive validity coefficients over time, though the term may also refer to any of several forms of change in job performance scores (e.g., Barrett, Caldwell, & Alexander, 1985). Unreliability, range restriction, and differential range restriction have been proposed as statistical artifacts that might explain the validity degradation problem (Barrett et al., 1985). Specifically, these artifacts would serve to attenuate validity coefficients, and may produce greater attenuation as time from selection increases.

Meta-analytic techniques have been used to correct the validity coefficients for these artifacts (Hulin et al., 1990). Hulin and colleagues found that 82% of the studies showed validity degradation, and corrected studies showed more extreme decay. Keil and Cortina (2001) also conducted a meta-analysis on dynamic criteria studies, using catastrophe theory to make predictions about the functional form of the decay, specifically, that validity decay would follow a cubic trend. These authors also found support for the ubiquity of validity degradation.

There are several explanations for why validity decays over time. The most commonly invoked are the *changing task* and *changing person* models (Alvares & Hulin, 1972; 1973). These models state that either tasks change over a period of time or practice such that the tasks draw on the abilities needed to perform them in a different way than they do when first learning the task, or that the person learns and changes on those attributes, with task demands remaining constant. Both approaches may be modeled using the formula:

$$Y_{ij} = \sum_{k=1}^K (a_{jk} \times x_{ijk}) + e_{ij} + s_j \quad (2.1)$$

where  $Y_{ij}$  is performance for the  $i$ th individual on the  $j$ th trial;  $a_{jk}$  is the weight of the  $k$ th common ability for determining performance on the  $j$ th trial;  $x_{ijk}$  is the  $i$ th person's score on the  $k$ th common ability at the time of

the  $j$ th trial;  $e_{ij}$  is the measurement error for person  $i$  on trial  $j$ ; and  $s_{ij}$  is a unique factor for trial  $j$ ; we assume the error terms have a mean of zero and are uncorrelated with each other and with scores on the common abilities. Therefore either a change in person scores,  $x_{ijk}$ , over trials (changing persons model) or change in ability weights,  $a_{jk}$ , over trials (changing task) could account for changes in performance.

Classical theories of skill acquisition would suggest that the changing person model does not make sense, as these models define abilities as relatively immutable attributes (e.g., Alvares & Hulin, 1972). These classical approaches make a sharp distinction between abilities and skills, where abilities are viewed as broad, general traits that affect a large number of tasks, and skills are viewed as narrower and more task-specific. Alvares and Hulin (1972) suggest that this approach is too strict a distinction (e.g., Humphreys, 1992). This sharp distinction is useful in some ways, as it creates a clear separation between distal predictors, proximal determinants, and performance (see section 2.7 for a quick overview). As a preview, the Campbell model (and related models, such as Motowidlo et al., 1997; Schmidt et al., 1986) holds that typical individual differences, such as cognitive abilities and personality traits, predict the presence and/or acquisition of knowledge, skill, and motivation/habits that are the direct causes of job performance behavior.

The model that will be tested in the current paper, however, is built around a changing persons interpretation. This is consistent with some research in the literature on dynamic criteria. For instance, Henry and Hulin (1987; 1989) argue that the changing task model would predict a plateau in validity degradation, which is not typically found (Henry & Hulin, 1987). In contrast to more classical theories of skill acquisition, the change in individuals is not just learning skills and changing abilities, but changing *identity* (Day et al., 2009; Day & Sin, 2009). As the individual is socialized to a particular work role, it is expected that the role identity will become more elaborated and, over a long enough period of time, may filter down to the individual's general self-concept (Roberts et al., 2006). This elaboration of identity will feed forward, and individuals who display characteristics consistent with the work role will develop beliefs and motivational structures that allow the acquisition of more job knowledge and skills and more effective work habits (Motowidlo et al., 1997).

For example, a police training program socializes individuals to the work

role of being a police officer, teaching them work skills and cultural norms for police officers. It is an environment likely to favor dominant individuals. For instance, a police recruit high in social dominance may be more comfortable using extreme physical control tactics, such as pepper spray or tasers. Successful use of these may foster greater efficacy beliefs for control tactics. These competence beliefs should, in turn, reinforce the recruit's identity as an effective police officer. Therefore, the individual becomes more comfortable with asserting control in difficult social situations and less reactive to the stress of those encounters. This self-assurance, in turn, may allow the individual to approach new challenges with mastery goal orientation, and attempt further exploration of his or her behavioral repertoire (cf., Day et al., 2009; Day & Sin, 2009; Roberts, 2006).

### 2.5.1 Multilevel perspectives on performance over time

Using a multilevel framework to analyze performance over time allows researchers to integrate two of Barrett, Caldwell, and Alexander's (1985) perspectives on dynamic criteria (Deadrick, Bennett, & Russell, 1997): (1) the decrease in validity over time, and (2) individual changes in performance over time. We disregard the third of Barrett et al.'s definitions of dynamic criteria, that of mean changes in performance, as it is largely irrelevant for the purposes of the argument presented here (cf., Austin, Humphreys, & Hulin, 1989). The first perspective is concerned primarily with validity degradation. This is the perspective that is generally most problematic for practitioners; the selection test works well for predicting initial performance, but its validity declines over time. The second perspective is concerned primarily with the instability of performance over time (e.g., Kane, 1986). Multilevel growth models allow the investigation of predictive validity for both performance intercepts and slopes, allowing for simultaneous inferences regarding initial performance and performance change. An additional strength of multilevel modeling is the ability to estimate both fixed effects and random effects, which represent general trends and individual trajectories, respectively (e.g., Singer, 1998).

The decay of validity over time should be of considerable concern to developers and users of selection systems. Echoing the definition of the ultimate

criterion, organizations are more interested in hiring those employees who will perform well over the course of their tenure, not just at the beginning of their employment. Of course, declining validity does not necessarily imply that any employees are becoming worse at their jobs, but it does mean that the tests used for selection provide less information about performance the further away in time that performance is measured. Thus, the prediction of performance over time becomes an issue. The identification of predictors of performance change may prove to be as important as predictors of early performance, if not more so (Hofmann, Jacobs, & Gerrass, 1993; Ployhart & Hakel, 1998).

Some hope for predicting performance trajectories may lie in personality trait measures; while the decline in validity for ability predictors is well-documented, the dynamic relationship of personality and performance is less studied. For instance, conscientiousness is generally positively related to job performance (e.g., Barrick & Mount, 1991), however, it has been found to negatively predict learning in training (Martocchio & Judge, 1997). Martocchio and Judge found that the negative impact of conscientiousness was mediated by the motivational variables self-efficacy and self-deception. However, other researchers have not found evidence for mediation (e.g., Lee & Klein, 2002). Lee and Klein also found that conscientiousness had no meaningful relation to early learning, but correlated .22 with later learning. As a whole, these results indicate that conscientiousness may be inconsequential to performance in the transition phase, when knowledge and skill are being acquired, but important during the maintenance phase, when those skills are well-practiced (Murphy, 1989).

In fact, conscientiousness has been found to display somewhat complex patterns in predicting performance trajectories (C. Thoresen, Bradley, Bliese, & D. Thoresen, 2004; Zyphur, Bradley, Landis, & C. Thoresen, 2008). In a comparison of quarterly sales performance between transitional and maintenance samples, Thoresen et al. (2004) found that both conscientiousness and extraversion predicted performance differences between participants, while only conscientiousness predicted performance growth in the maintenance sample. Agreeableness and openness to experience predicted both performance differences and trends in the transition phase. Zyphur and colleagues (2008) used a censored latent growth model to predict seven semesters of college GPA using cognitive ability and personality. Their findings showed that



both cognitive ability and conscientiousness predicted initial performance, with cognitive ability the stronger predictor. However, conscientiousness positively predicted performance trajectories, when controlling for initial performance, and became the better predictor after the third semester.

The primary practical implication of dynamic criteria is validity degradation. Furthermore, to the extent that it is not clear what the sources of stability and change are in dynamic criteria, a scientific understanding of the performance domain is hampered. The ubiquitous decline in validity for cognitive predictors adds to the call for more holistic pre-employment assessment. Other predictors may better explain performance in the long term (e.g., in the maintenance phase; Murphy, 1989). Identifying individual characteristics that differentially predict performance intercepts and slopes would be useful in designing selection systems with long-term utility. Furthermore, a solid scientific theory of the development of the performance repertoire may allow insights into how to accelerate the process, which would assist in developing training and development programs. In the next section, a framework for validating selection instruments using multivariate dynamic criteria is developed.

## 2.6 Validation: A causal model of job performance

A central task for industrial and organizational psychologists has always been the prediction of job performance. This is also a central concern for business organizations. Quality measures of individual differences that strongly predict job performance allow for better selection systems and hiring decisions. It is critical to acknowledge that validity is not a property of an instrument itself. Validity refers instead to the quality of inferences made about scores on that instrument (Angoff, 1988). It is not the case that a written ability test is somehow inherently valid, but that scores on the measure reliably predict performance on the job; the test, therefore, can validly be used to select applicants for a position.

When an applied psychologist designs a predictive validity study, that psychologist is, at least implicitly, making a causal hypothesis. Scores on the selection instrument are the trace evidence of a psychological phenomenon, such as the tested individual's cognitive ability. Criterion measures are simi-

larly trace evidence of the job performance behavioral repertoire. The point of validation is to ensure that these measures reflect evidence of those constructs, and there is a meaningful predictive relationship between the constructs (Binning & Barrett, 1989; cf., Schmidt et al., 1986). A claim that a test has predictive validity is really a claim that the phenomenon of cognitive ability bears some causal connection to the phenomenon of job performance. Various studies have tested “causal models” for the cognitive ability to job performance relationship (Borman et al., 1991; Hunter, 1983; Schmidt et al., 1986; Schmidt & Hunter, 1992). These models all demonstrate that the influences of distal traits are mediated by the individual’s current repertoire of knowledge and skill (cf., Humphreys, 1992).

Validation research, from the standpoint of personnel selection, is fundamentally a search for the causal model underlying job performance. In order to craft this model, we must first collect information on appropriate criteria, which are likely multidimensional. We must then establish what are the direct antecedents of performance facets, for example declarative knowledge, procedural knowledge and skill, and work habits or motivation (Campbell, 1990; Motowidlo et al., 1997). We can then assess whether the data are consistent with a causal model with pathways from deep-seated traits, such as cognitive ability and personality traits, through these knowledge/skill and habit/motivation constructs to performance criteria (Borman et al., 1991; Hunter, 1983; Schmidt et al., 1986). These models have been tested with each of the main variables measured once (though not necessarily at the same time). A process approach, with variables sampled at appropriate and multiple time points based on adult development theory, would provide stronger evidence for the process by which an individual’s personal resources become invested in knowledge, skill, and habits, which then allow the individual to engage in performance behaviors.

## 2.7 Validation evidence: A review

Next, the current evidence on the validity of cognitive ability and personality measures will be examined. Then, evidence for the causal processes by which cognitive ability and personality influence behaviors in the performance repertoire will be reviewed. The logic behind establishing such

causal models is that understanding the linkages between distal individual differences and work behavior is enriched by understanding the more proximal determinants of behavior more accurately. Such knowledge can help researchers to specify the circumstances in which a particular trait may be useful for predicting a particular facet of performance over a given time period. For example, understanding how goal constructs may mediate the influence of personality traits on behavior helps to clarify how the social setting of work may moderate the validity of certain personality traits, such as extraversion and agreeableness (Barrick, Mount, & Stewart, 2003).

### 2.7.1 Cognitive Ability

Cognitive ability has been demonstrated repeatedly as the single best predictor of job performance (Ree & Earles, 1992, Hunter & Hunter, 1984, Schmidt & Hunter, 1998). Meta-analytic estimates of the correlation between cognitive ability tests and performance criteria are about .51 (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). Only work sample tests demonstrate higher validity (.54; Schmidt & Hunter, 1998). However, these tests require some job knowledge, making them inappropriate to entry-level jobs, and these tests are relatively more costly than written cognitive ability tests. Additionally, work samples may more appropriately be placed in the portion of the nomological network of job performance occupied by job knowledge and skill. Additionally, cognitive ability tests are generally effective at predicting both academic and job success, to approximately the same degree (Kuncel, Hezlett, & Ones, 2004).

Why are cognitive ability tests so effective at predicting job performance? The Campbell model (Campbell, 1990; Campbell et al., 1993) posits that performance is directly determined by an individual's declarative knowledge, procedural knowledge, and motivation. These, in turn, are partially determined by individual differences in cognitive ability and personality traits, such as conscientiousness. The implication here is that cognitive ability is an indirect cause of job performance, its effect being mediated through the acquisition of job knowledge and skill. Support for this contention has been found (e.g., Hunter, 1986; Schmidt & Hunter, 1992; Schmidt et al., 1986). The mediation by job knowledge would make tests of such knowledge more

proximal predictors of job performance, though not always appropriate as knowledge and skill may be acquired on the job.

There may, however, be some problems with this accumulated evidence. The definition of performance is often “notoriously vague” (Murphy & Shiarella, 1997, p. 846). Furthermore, as the world of work continues to change, for example as individual task performance becomes more and more nested within team contexts, the facets of performance that are most strongly related to cognitive ability may decrease in importance to many organizations.

### 2.7.2 Personality

One personality trait that has been found to have an indirect effect is conscientiousness, similarly mediated by the direct determinant, motivation (Campbell, Gasser, & Oswald, 1996). According to meta-analytic estimates, the greatest total validity for two predictors is obtained by administering a cognitive ability test with an integrity test ( $R = .65$ ; Schmidt & Hunter, 1998). This may be partially explained by the finding that integrity tests tap “all that is good” from several useful personality constructs (conscientiousness, agreeableness, and emotional stability; Ones, 1993). Integrity tests also show loadings on a general factor, indicating that they may tap a broader construct than just amalgamating the three above mentioned personality traits (Ones, 1993).

The incremental validity of integrity tests, and their relationship to personality variables leads to the question, how good are personality variables at predicting job performance? Depending on the personality construct being measured and the care with which it is matched to criterion measures, the prediction is quite good, in fact (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991; Hogan & Holland, 2003). Conscientiousness, often described in terms of dependability and achievement-orientation, is found to be predictive across many jobs, with a meta-analytic corrected estimate of validity in the area of .22 (Barrick & Mount, 1991). When matched with criteria theoretically related to the personality construct under study, the Hogan Personality Inventory shows strong predictive validity estimates, with emotional stability showing the highest validity, in the realm of .40 (Hogan & Holland, 2003).

These constructs may be tied to job performance through their influence on direct determinants, such as motivation or work habits.<sup>1</sup> For instance, conscientiousness is related to motivational constructs (Judge & Ilies, 2002), such as goal choice (Gellatly, 1996), goal commitment (Hollenbeck, Williams, & Klein, 1989), and generalized self-efficacy (Judge, Erez, Bono, & Thoreson, 2002). Motivation in the Campbell model is defined as the choice to perform, the level of effort to demonstrate in performance, and to persevere in performance over time, which is similar to some views of conscientiousness.

The primary difference is that conscientiousness is seen as a relatively stable trait, while motivation is viewed as more state-like (Kanfer & Ackerman, 1989). There is evidence demonstrating the link between conscientiousness and motivation in training, with a correlation of .38 (Colquitt, LePine, & Noe, 2000). Thus, motivation may be seen as the most proximal non-cognitive variable in predicting performance, but assessing such a state variable for personnel selection is not a good option. In line with this reasoning, personality traits such as conscientiousness can be viewed as a propensity distribution of engagement in behaviors in the domain defined by that construct (cf. Fleeson, 2001). From this perspective, conscientiousness measures may predict performance by their ability to predict the relative likelihood that individuals will be engaging in dependable, reliable and goal-oriented behaviors. Additionally, Barrick, Mount, and Strauss (1993) provide evidence that the influence of conscientiousness on supervisory ratings of performance is partially mediated by self-set goals. Individuals high in conscientiousness are more likely to select goals and be more committed to their goals; however, conscientiousness maintains a direct relationship to performance ratings, indicating other pathways for its influence.

Conscientiousness has received a great deal of the attention in the I/O psychology literature, perhaps because Barrick and Mount's (1991) meta-analysis found it to be a valid predictor for all of their occupational groups and performance criteria. But conscientiousness is not the only personality trait that is predictive of performance. Using a meta-analytic sample of confirmatory studies, where a particular personality trait had been *a pri-*

---

<sup>1</sup>It is possible that personality could also contribute to the development of job knowledge and skill, for instance a conscientious worker may invest more heavily in study and practice. Such a link is not often included in causal models of performance, such as Motowidlo, Borman, and Schmit's (1997) approach, but this does not rule out the possibility.

*ori* hypothesized to predict an outcome, Tett and colleagues (1991) found that both Agreeableness and Openness to Experience provided reasonable validity coefficients (but see Ones, Mount, Barrick, & Hunter, 1994 for a critique of the findings and methods of this study). In a meta-analysis focusing on the Hogan Personality Inventory, and matching personality trait measures with performance criteria built to reflect behaviors relevant to particular traits, Hogan and Holland (1993) found Emotional Stability to be a generalizable predictor of job performance. Furthermore, it is expected that the relationship between any personality trait and any performance criterion is moderated by both job autonomy and job type. For instance, Extraversion is likely to be a more valid predictor of performance for salespersons, whose jobs require a great deal of interaction with other people, than for bookkeepers, whose jobs require significantly less human interaction. Autonomy may influence the strength of the relationships between personality traits and criteria, in that low autonomy jobs allow relatively little freedom of behavioral expression, whereas high autonomy jobs do. Therefore, personality trait measures should be most predictive in high autonomy jobs (e.g., Barrick, Mount, & Strauss, 1993).

## 2.8 Lifespan development perspectives on the performance repertoire

The validation evidence discussed above represents the beginning of a process model for the development of the performance repertoire. One of the purposes of this paper is to develop this process model more thoroughly. Most approaches to the question of dynamic criteria have been explicitly cognitive (e.g., Ackerman, 1987; Murphy, 1989; Zyphur et al., 2008), or have used early performance as a proxy for ability (e.g., Henry & Hulin, 1987). The approach advocated here will place the dynamic criteria problem within the broader adult development approach, which includes personality characteristics in development (e.g., Roberts, 2006).

The performance development process is conceptualized in a multilevel fashion: the discrete performance behaviors are a surface-level manifestation of the individual's attributes at the lower levels. Beneath these surface behaviors are self-regulatory mechanisms that direct attention and effort to-

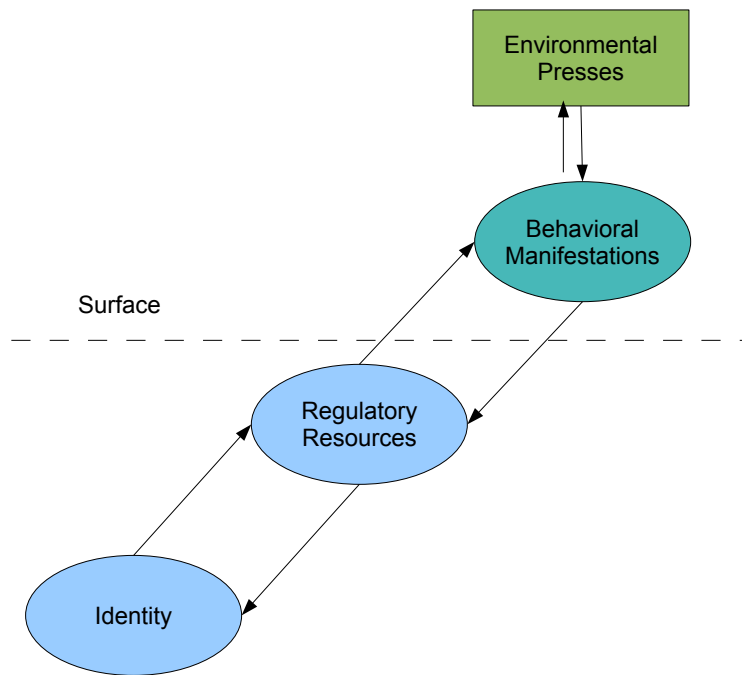
ward particular goals, which may or may not be relevant to the organization (Beal et al., 2005; DeShon & Gillespie, 2005). At the lowest level are identity mechanisms, deep-seated conceptions of the self which may serve to choose goals and actions (e.g., Hogan, Hogan, & Roberts, 1996). Each of these levels is viewed as dynamic and developmental; however, the timescales involved in each level's change process may differ. For instance, we would expect that behavior change can be relatively rapid, with regulatory changes being similarly fluid, while change in identity may require greater personal investment, and may therefore take significantly greater lengths of time (cf. Helson, Kwan, John, & Jones, 2002). Figure 2.2 shows a simple schematic representation of the developmental model that will be developed throughout this section.

As can be seen in Figure 2.2, the individual's identity is the foundation on which all other levels rest. How the individual views him- or herself will serve to help shape schematic knowledge structures, enforce particular viewpoints in top-down processing, direct attention in characteristic ways, and help shape personal projects, attachment patterns, and goals (Roberts, 2006). In turn, these social-cognitive mechanisms result in particular affective and regulatory manifestations during any given behavioral episode (Beal et al., 2005). These affective and regulatory mechanisms will interact with particular environmental demands to produce behavioral responses. Note that all levels are *dynamically* connected. It is expected that adjacent levels may mutually influence one another.

### 2.8.1 Identity and self-regulation: Mechanisms underlying behavior in performance episodes

The hierarchy of traits described above meshes with the Neo-Socioanalytic Model of personality, which posits that individual differences are arranged in several hierarchical domains, including cognitive abilities, personality traits, values, and motives (Roberts, 2006). In particular, traits and motives have long been alternative viewpoints to the organization of personality (Winter, John, Stewart, Klohen, & Duncan, 1998). The constructs at the highest level of organization are very broad and generalize over most specific situational contexts. This does not mean these broad constructs are “decontextualized”, or in some way context-free, but that they are built out of and define themes

Figure 2.2: Schematic Representation of a Developmental Model of Performance.



underlying an individual's characteristic way of interacting with the social world.

The general consensus is that something akin to the Big Five (or Six or Seven) surmounts the trait hierarchy and that broad life goals lie at the top of the motive hierarchy (Roberts, 2006). These two domains are conceptually distinct, but related (Roberts & Robins, 2000). The lower, more situationally-contextual levels of the hierarchy are believed to mediate the influence of the higher-level constructs on behavior (Bogg, Voss, Wood, & Roberts, 2008). While this section describes constructs such as personality traits as being at the top of their respective hierarchies, this means that they



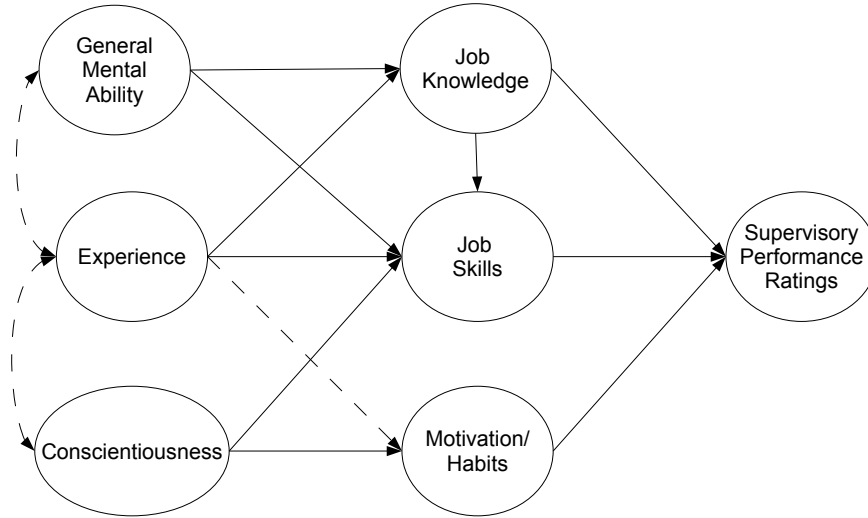
are at a deeper level; these constructs are highly identity-relevant and generalize over the specific roles that the individual may be expected to inhabit.

This pattern of mediation of deeper traits by more contextual traits is consistent with causal models of job performance (e.g., Schmidt & Hunter, 1992; Schmidt et al., 1986). Figure 2.3 presents a conceptual representation of such a basic causal model. Specifically, cognitive ability is not expected to directly cause job performance. An individual's level of cognitive ability instead influences the acquisition of job-specific knowledge and skill. The knowledge and skills, in turn, are viewed as the proximal determinants of job performance behaviors (Campbell, 1990). Studies such as Hunter (1983) and Schmidt et al. (1986) include indicators of job experience as a proxy for developmental opportunity. These models usually allow direct paths from both ability and experience onto knowledge and skill, and allow knowledge to affect skill.

Similar patterns are expected in the non-cognitive domain, such that social skills may be more relevant for contextual performance (Motowidlo et al., 1997), though the evidence for this is more indirect (Gellatly, 1996). The model presented in Figure 2.3 allows personality (shown only as conscientiousness) to influence work skills and work habits/motivation, and for experience to influence work habits. This model is not put forth as a specific set of predictions, but as a schematic representation of the adult development-performance repertoire development process.

Let us return to the neo-socianalytic model (Roberts, 2006). We take the example of an individual beginning his or her first "real" job. We expect the environment, as represented by the role demands and expectations of the job, to press for effortful and goal-directed behavior, behaviors within the general domain of conscientiousness. We would expect that over time, the investment the individual puts into this work role would perhaps lead him or her to view him- or herself as a "hard worker". This change is a change in the individual's role identity. The person may not yet view him- or herself as an "industrious" individual, however. Such a self-judgment would reflect change in their more general identity (Wood & Roberts, 2006). As time passes, and the individual becomes more deeply embedded within their job, we would expect greater investment in the work role, which could trickle down to the deeper general identity, leading to a self-view of being "industrious" and "conscientious". This example illustrates the general argument made

Figure 2.3: Basic causal model of performance ratings



in this thesis. As individuals adjust to a work role, they will craft their behaviors such that their personal repertoires are adaptive for their respective environments (Kerr, 1995) and their personal characteristics (Roberts, 2006). The effectiveness of this crafting process provides feedback to the individual on their competence and appropriateness in the role. This feedback affects self-views, affecting general and role-specific identities. These revised self-views, in turn, affect future behavior patterns.

Over a variety of individuals in a variety of contexts, such social investment processes are likely. The study described herein takes place within a particular context, a police training academy. Individuals have self-selected, and in turn, have been selected into this environment. As part of the the job analysis for this project, police officers described themselves as individuals who “like to carry a gun and drive fast”, and their jobs as “long stretches of boredom, with moments of sheer terror”. Many individuals attempting to become police officers have previous work experience, often having served in a branch of the armed forces or in some other form of protective service, such as emergency medical technicians, police dispatchers, or corrections officers. Therefore, it is likely that individuals who attempt to become police officers

have a reasonable expectation of what the job is like.

Further, developmental experiences often follow the “corresponsive principle” (Roberts & Caspi, 2003). For instance, individuals who are somewhat impulsive, dominant, and low in stress reactivity may choose to become police officers. Training and on-the-job experiences would, in turn, favor these traits which results in deepening and elaboration of these dispositions. In the present study, the training academy is viewed as a socializing experience that favors displays of social dominance and low reactivity to stress. For instance, as part of training in “control tactics” recruits are expected to fire a taser at another recruit, which is a dominant act, and to be fired upon, which is a stressful experience. Therefore, it is expected that traits related to dominance, particularly social potency will increase, and that traits related to anxiety and stress reactivity, particularly neuroticism and stress reactance will decrease.

## 2.8.2 Building a framework for the validation of multivariate dynamic criteria

Criteria measurements representing the job performance repertoire are best conceptualized as multivariate and dynamic. Murphy has suggested methods for dealing with multivariate criteria (Murphy, 2009; in press; Murphy & Shiarella, 1997). Murphy’s discussion limits the dimensionality of performance to task and contextual performance, but provides other criteria of interest, such as the likelihood that using a particular test may produce adverse impact. As mentioned above, these methods produce a single index of validity, similar to standard validation procedures. However, the methods allow an organizational decision-maker to specify the importance of a particular criterion. In addition to these multivariate methods, there are several approaches to modeling the time factor in performance measurements. This includes the multilevel approach taken by Deadrick et al. (1997) and Hofmann et al. (1993), latent growth approaches (e.g., Chan, 1998), and autoregressive latent trajectory models (Zyphur et al., 2007). All of these approaches suffer from univariate thinking, although Chan’s approach does allow multiple indicators.

First, consider why scientists and practitioners should care about both

the dimensional structure and the temporal dynamics of performance in designing validity studies. Dimensions of performance are both definitionally and statistically separable. These dimensions may show differential patterns of stability and change over time, which may be obscured by condensing them into a single dimension. Recall the example regarding Extraversion and its facets, Social Vitality and Social Dominance. These facets display different patterns of growth and decline over the life course, with social dominance increasing with age and social vitality decreasing (Roberts, Walton, & Veichtbauer, 2006). If one were to examine only aggregate extraversion, these patterns of change would be hidden.

This suggests that individuals may engage in differential investment in particular performance domains. For example, an individual experiencing consistent negative environmental feedback regarding his task performance may choose to invest more effort into building social capital in the workplace, offering what assistance he can and generally being a pleasant coworker. This could serve as a compensation mechanism, where the individual realizes that the goal to be a highly competent worker is impossible and is replaced by the goal to achieve value via social rather than technical means. In this case, the individual's trajectory for citizenship performance may be positive, while his trajectory for focal performance may remain flat or even display a negative trend. Therefore, even if performance facets can agglomerate into a general factor at any measurement opportunity, to use such a gross, aggregate measure may obscure different trajectories for different performance facets. It is also possible that individuals who do not fit with the demands and reward structures of the environment may simply choose to leave (Roberts, 2006; Schneider, 1987)

## 2.9 The current study

### 2.9.1 Research questions

The primary question asked in this study is: how do cognitive abilities and personality traits jointly predict performance on multiple police work criteria over time? Additionally, how do personality traits change in response to the environmental press of learning a high stress job? Furthermore, the study

addresses how changes in motivational structures influence performance on different criteria. In essence, the study addresses how to effectively track the multidimensional and dynamic properties of both sides of the performance prediction problem. These research questions coalesce in an attempt to validate a selection battery by developing a process model of multidimensional dynamic performance criteria.

To address these questions, specific hypotheses are outlined below.

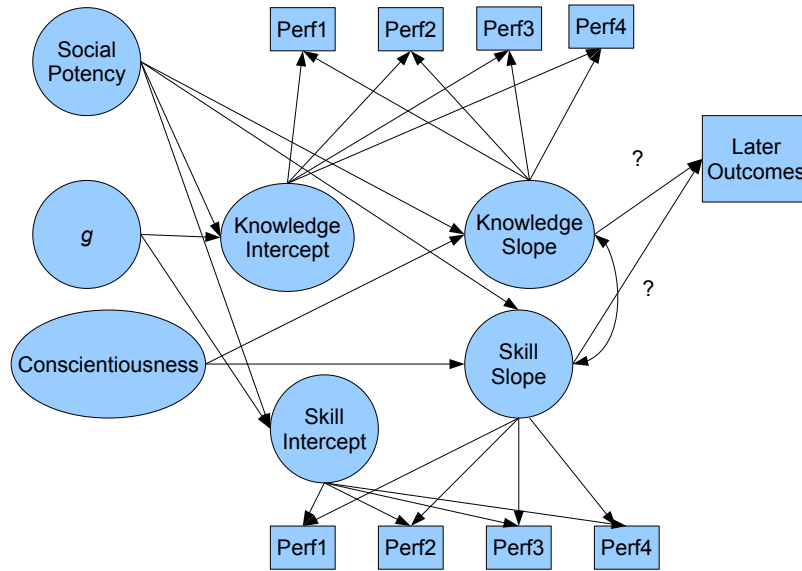
### 2.9.2 The causal model

The outline of this study follows the logic of the basic causal model of Figure 2.3. Individual differences measures, such as tests of cognitive ability or personality traits, can serve as selection measures before applicants are hired. Then knowledge and skill repertoires can be measured during training or on-the-job, and final supervisory ratings or other indices of job performance can be collected. These are the approaches used in most of the previous path analytic studies (e.g., Schmidt et al., 1986).

The model in Figure 2.3 suffers from one major flaw. Work experience is expected to influence the job knowledge and skills constructs as more time on the job allows greater knowledge and skill acquisition. Experience is therefore a proxy indicator of developmental opportunity. This study did not include a measure of experience, because all participants are trainees with no direct experience of the job. Instead, repeated measures of job knowledge and work skills were taken. Job knowledge was indexed by 4 written examinations given throughout the training program. Work skills were indexed by instructor ratings of several behavioral demonstration exercises: simulated vehicle stops, simulated response to disturbance calls (integrated scenarios), firearms, and control tactics, taken at weeks 3, 6, 10, and 12. Measures of cognitive ability and personality were taken on the first day of training, and are consistent with the selection mechanisms typically used for police officers.

To capture changes in the participants' identities, the personality measures and motives for police work were given at each followup assessment (weeks 3, 6, 9, and 12), around the times that police recruits took their four exams. Changes in regulatory mechanisms were indexed by two indicators. Regulatory mechanisms were measured using job-specific self-efficacy (Jones,

Figure 2.4: Example individual growth model of performance ratings



1986) and work goal orientation (Vandewalle, 1997).

The logic of the basic causal model examined in this study is presented in Figure 2.4.

### 2.9.3 Hypotheses

In general, to understand the relationships between broad constructs, their more fine-grained aspects need to be measured and modeled (Schneider, Hough, & Dunnette, 1996). Dalal and Hulin (2008) have called for a multivariate dynamic approach for criterion measurement. They suggest using multivariate dynamic criteria in an attempt to answer the question, “motivation for what?” This is particularly relevant to the questions addressed in this study, as we are attempting to map changes in performance criteria onto changes in the determinants of performance behaviors (e.g., Bashshur, 2006). Indeed, the primary expectations of this study were that different facets of job performance would be found, and that these facets would display different patterns of change over time, because their relationships with the underlying regulatory and identity mechanisms differ from one another.

Multilevel models were used to investigate the prediction of dynamic performance. These models allow the prediction of initial performance and changes in performance over time. Again, as suggested by Murphy (1989), it is expected that cognitive ability will predict initial performance best, with declining validity. Hypotheses tested using multilevel modeling are framed in terms of both “fixed effects”, or the average effect across participants, and “random effects”, or individual variability about these means.

It is expected that personality traits, in particular conscientiousness, will predict more weakly to begin with, but will actually increase in predictive validity.

*H1:* Evidence will be found for four dimensions of performance, consisting of academic performance, physical performance (those aspects of police work related to movement and dexterity), operations performance (unique aspects of police work, such as patrol operations and vehicle stops), and contextual performance (particularly dependability and maintaining good human relations).

*H2:* The pattern of change will be unique to each facet of job performance. Specifically, the intercept and slope of time will differ for each facet.

The implications of this hypothesis is that the fixed effects should differ between performance facets. Since the study takes place in a training context, it is expected that all program-relevant aspects of performance should show positive growth trajectories. It is expected that academic performance, which is little different from a school setting should show the smallest slope and that operational performance, which is essentially unique to police settings, should show the steepest slope. Physical performance is likely to fall between these extremes. Contextual performance is unlikely to show any particular fixed effect.

Individuals differ in various personal characteristics starting a training program. Therefore, there should be individual differences in performance trajectories (e.g., Day & Sin, 2009), particularly intercepts. Further, it is likely that individuals do not develop at the same pace, due in part to differences in those individual characteristics, which would result in differences

in slopes between individuals.

*H3*: Individuals will display unique performance trajectories.

Hypothesis 3 would be supported by significant random effects on the performance facets' intercepts and slopes (i.e., significant inter-individual variance in growth curve intercepts and slopes).

Based on the theorizing above derived from Motivated Action Theory (DeShon & Gillespie, 2005) and adult development (Baltes, 1997; Day et al., 2009), these individual trajectories should be driven by individual differences in the development of the employee's identity and self-regulatory skills.

*H4*: The relationship between performance and self-regulatory mechanisms and identity mechanisms will be unique for each performance facet. Specifically, the performance facets will have different regression coefficients for each within-person (time-varying) regulatory covariate (Goal Orientation and Self-Efficacy) and for each within-person (time-varying) identity covariate (Social Potency and Stress Reactance).

Multivariate validation research suggests that individual difference predictors may differentially predict different performance facets (e.g., Murphy & Shiarella, 1997). For instance, conscientiousness is a better predictor of contextual performance than it is of task performance (Borman & Motowidlo, 1993). Additionally, prior evidence on dynamic criteria suggests that personality should better predict performance slopes than cognitive ability (Zyphur et al, 2007).

*H5*: A model with unique regression coefficients from each performance facet on individual difference predictors will outperform a model where those regression coefficients are constrained to equality.

*H6a*: Personality predictors and cognitive ability will both predict intercepts for the individual growth curves.

*H6b*: Personality predictors will demonstrate stronger relationships with



performance slopes on time than will cognitive ability.

#### 2.9.4 Ancillary predictions

Hypotheses 4 and 5 are fairly general. They do not constitute strong tests of the underlying model put forth in this paper. Adding slight refinements, however, allow stronger tests. The causal model to be tested indicates that those identity elements relevant to the work role should change in response to the role demands. Therefore, those traits most strongly related with the recruits' identities as police officers should be positively associated with change in performance, while those that have little to do with their identities should not. We suggest that changes in Social Potency should be positively related, and Stress Reactance negatively related to changes in job performance facets. We suggest this because a police training program is essentially an effort to increase recruits' capabilities in handling and commanding complex, difficult, and intense social situations. We also believe that changes in Affective Identity motives should positively relate to changes in performance facets because, as recruits develop a clearer and stronger identity as a police officer, these motives should matter more (Lodi-Smith & Roberts, 2007). This argument rests on the corresponive principle, so those traits that lead individuals to become police officers should change more than unrelated traits (Roberts & Caspi, 2003). We suggest that conscientiousness traits should actually not be highly related to performance, as conscientiousness seems to be less important for law enforcement than for many other jobs (Barrett, Miguel, Hurd, Lueke, & Tan, 2003).

The current study aimed to address the questions raised above regarding the conduct of validation studies in the face of performance criteria that are multidimensional and dynamic. First, the validity of cognitive ability and personality trait measures for predicting several performance outcomes over time was assessed. The this is modeled as a process of individual development, with change in personality and self-regulatory mechanisms influencing changes in performance.

## Chapter 3

# STUDY OVERVIEW

### 3.1 Sample

The current study uses a special sample: police recruits were measured on the first day of their training academy. The police recruits were drawn from many police departments in northern and central Illinois, excluding the city of Chicago. Their performance was measured with a variety of objective measures, examination scores, instructor ratings, and self-evaluations throughout the 12 week program. We do not have a strong theory for the development of performance in this setting. Performance was naturally assessed during the training program quarterly, broken up by 4 examinations. The self-evaluations were taken at four times, every three weeks during the program. These self-evaluations serve as the primary basis for examining developmental trends in several performance facets. It is expected that this the first day of this training program represents as much of a true time zero as can be found in a real work setting. Therefore, a sample like this will help to build theory that may aid future researchers in determining the proper timing for criterion measurements.

The context of this sample is important. The participants are mostly in their late 20s. Many have already served in the military, or have worked in some form of law enforcement setting in the past, such as dispatch or corrections. Therefore, the participants have some previous work experience, and some familiarity with the culture and demands of police work. It is likely that the sample is not representative of the general population. They are somewhat older than many individuals beginning their careers. They are also likely more socially dominant, more impulsive, and less anxious than typical.

## 3.2 Procedure

The study was introduced to the trainees during the Sunday evening orientation to the training program. On the first day of training, recruits who elected to participate in the study were given the individual differences measures. Exams were given during weeks 3, 6, 10, and 12. Instructors provided ratings on several performance measures at the same time. Because of this timing, followup surveys of personality traits, motivation, and self-evaluations of performance were given quarterly (weeks 3, 6, 9, and 12).

## 3.3 Measures

Cognitive ability was assessed using the Wonderlic Personnel Test (Wonderlic, Inc, 1999). Only total scores were recorded, which makes a local estimate of reliability impossible. The norm sample alpha for the Wonderlic is approximately .88 (Wonderlic, Inc, 1999). Personality was assessed using the Big Five Inventory (John & Srivastava, 1999) and the Social Potency and Stress Reactance scales of the Iowa Personality Inventory (Donnellen, Conger, & Burzette, 2005), which is a short form of the Multidimensional Personality Questionnaire (Tellegen, 1982). Motivational variables measured were Work Goal Orientation (Vandewalle, 1997), and Job-specific Self-efficacy (Jones, 1986). The alpha reliabilities for each measurement occasion for each of these scales is provided below. Alphas for the Big Five Inventory on the first day of training are reported in Table 5.2.

Reliability estimates for the Big Five Inventory were as follows. During week 3, extraversion showed an alpha of .86, agreeableness (.80), conscientiousness (.85), neuroticism (.81), and openness (.79). For week 6, extraversion had an alpha of .85, agreeableness (.81), conscientiousness (.86), neuroticism (.84), and openness (.82). During week 9, extraversion had an alpha of .86, agreeableness (.83), conscientiousness (.88), neuroticism (.86), and openness (.84). For week 12, extraversion had an alpha of .86, agreeableness (.84), conscientiousness (.89), neuroticism (.85), and openness (.84).

Reliability estimates for the Iowa Personality inventory were as follows. At week 3, social potency had an alpha of .67, and stress reactance had an alpha of .76. For week 6, social potency had an alpha of .61, and stress

reactance had an alpha of .83. During week 9, social potency had an alpha of .67, and stress reactance had an alpha of .67. During the final week, social potency showed an alpha of .74, and stress reactance had an alpha of .78.

The reliability estimates for each of the motivation scales follow. At week 3, learning goal orientation had an alpha of .85, performance-prove (.76), performance-avoid (.82), and self-efficacy (.74). For week 6, learning goal orientation had an alpha of .88, performance-prove (.86), performance-avoid go (.84), and self-efficacy (.76). For week 9, learning goal orientation had an alpha of .88, performance-prove (.85), performance-avoid (.83), and alpha for self-efficacy was .75. For the last measurement occasion, alpha for learning goal orientation was .86, performance-prove was .89, performance-avoid go was .82, and self-efficacy was .74.

A variety of training outcomes served as criteria. Scores for four written examinations functioned learning criteria and the pass/fail rating on the state proficiency exam served as a knowledge retention criterion. Self-ratings of training engagement, punctuality, citizenship performance, maintenance of personal discipline and performance in simulation exercises were collected every three weeks during the twelve week training program. Instructor ratings on these same scales were collected during the final week of training, with ratings of performance in the simulation exercises and firearms and control tactics collected contemporaneously with the self-reports. Hours absent and awards received serve as distal criteria.

The design captured as between-person predictors standard selection measures (time-invariant covariates). Change on a variety of identity and self-regulation mechanisms were taken via self-report every three weeks, and served as within-person covariates (time-varying covariates). Time-varying dependent measures were collected by self-report along with identity and self-regulation measures. The four examinations occurred at about the same time as these measurements, but should not be considered as occurring simultaneously with them. The same goes for instructor ratings of the performance domains, which were collected after each exam.

## Chapter 4

# ANALYTICAL STRATEGY

### 4.1 Static hypotheses

Hypotheses regarding the structure of performance were addressed using a mix of confirmatory factor analysis and exploratory item clustering techniques. The *a priori* expectation was that a four factor structure would be found for performance in training, consisting of citizenship, academic, physical, and operational performance facets. Since previous research has found general factors in performance (Hulin, 1982; Viswesvaran et al., 2005), this four factor structure was compared to a model with only a general factor and to a bifactor model, with both four primary factor and a general factor influences performance indicators. Additionally, the facet structure of performance were examined using Revelle’s (1979) item cluster analysis.

### 4.2 Dynamic hypotheses

Dynamic hypotheses were tested using multilevel modeling approaches, particularly random coefficients modeling (e.g., Bliese & Ployhart, 2002), implemented using SAS 9.1 PROC MIXED (e.g., Bryk & Raudenbush, 1992; Singer, 1998). Similar to Deadrick et al.’s (1997) work, these analyses allow modeling of both performance slopes and intercepts. The general form of the models tested is presented below. Note that this is a three-level model, with level one being the dependent variable, level 2 being within person (i.e., over time), and level 3 being between persons. A SAS programming trick allows estimation of levels 1 and 2 as a single level (Equation 4.4), but gives an estimate of the covariance between the different dependent variables. This is accomplished by stacking the dependent variables into one Y-term, and using

a dummy variable to indicate which variable is in use (Hoffman & Rovine, 2007). For two dependent variables, the model appears as:

$$\text{Level 1 : } Y_{tid} = B_{0i1}(DV1) + B_{1i2}(DV2) \quad (4.1)$$

$$\text{Level 2 : } B_{0i1} = \delta_{0i1} + \delta_{1i1}time_{ti1} + e_{ti1} \quad (4.2)$$

$$B_{1i2} = \delta_{0i2} + \delta_{1i2}time_{ti2} + e_{ti2}$$

$$\text{Level 3 : } \delta_{0i1} = \gamma_{001} + \gamma_{011}X_1 + U_{0i1} \quad (4.3)$$

$$\delta_{1i1} = \gamma_{101} + \gamma_{111}X_1 + U_{1i1}$$

$$\delta_{0i2} = \gamma_{002} + \gamma_{012}X_1 + U_{0i2}$$

$$\delta_{1i2} = \gamma_{102} + \gamma_{112}X_1 + U_{1i2}$$

where the data are structured such that the  $T$  measurements for  $I$  participants on  $D$  outcomes are stacked on top of each other: participant is the slowest running index, with outcome being the next slowest, and measurement occasion being the fastest running index (i.e., for participant 1, we have  $T$  measurements on DV1, then  $T$  measurements on DV2). Therefore, the set of outcome measures are held in a  $TID \times 1$  vector:  $Y_{tid}$  indicates the  $t^{th}$  measurement for participant  $i$  on performance facet  $d$ .

The DV variables are dummy codes that indicate which performance dimension is currently being modeled.  $Time_{ti}$  is a linear index of time and serves as the basis vector for a linear growth trend. Terms for  $Time_{ti}^2$  or higher-order effects can be added to test for curvature in the growth patterns. The  $\delta_{0id}$  and  $\delta_{1id}$  terms represent the intercept and slopes for the performance growth model. The  $X_n$  terms are time-invariant covariates representing a participant's standing on some individual difference measure, such as the Wonderlic Personnel Test or conscientiousness. The  $e_{tid}$  and the  $U$  terms are errors at the within- and between-persons levels, respectively (e.g., Hofmann, Jacobs, & Baratta, 1993). The  $U$  terms allow for individual variability in the level-2  $\delta$  parameters (i.e., the variance of these terms are the random effects).

Additionally, time-varying covariates can be modeled by adding additional  $\delta$  terms to equation 4.2. This allows tests for hypotheses regarding how standing on a particular regulatory or identity variable, such as Self-efficacy or Stress Reactance, at a given measurement occasion may influence concurrent performance across the performance facets.

It is possible to rewrite Equations 4.1 and 4.2 as one level, collapsing levels 1 and 2 into a single level; the level 2 regression equation in the parentheses is estimated for each DV, allowing covariances between DVs to be estimated by using their dummy codes, i.e., the DV dummy “wakes up” the regression for a specific dependent variable,  $d$ :

$$Y_{tid} = DV1(\delta_{0i1} + \delta_{1i1}time_{ti} + e_{ti1}) + DV2(\delta_{0i2} + \delta_{1i2}time_{ti} + e_{ti2}) \quad (4.4)$$

This model allows for inference tests regarding the competing models by way of likelihood ratios. For instance, to determine whether there are unique developmental trends in the facets of performance, we may freely estimate  $\delta_{1id}$  for each  $DV$  or constrain them to be equal. For four performance facets, this likelihood ratio is distributed as a  $\chi^2$  with  $df = 3$ . We can constrain the coefficients for time or time-varying covariates to be equal across DVs by altering Equation 4.4 as:

$$Y_{tid} = \sum_{d=1}^D DV_d(\delta_{0id} + e_{tid}) + \delta_{1i}time_{ti} \quad (4.5)$$

such that  $\delta_{1i}$  is “awake” for all DVs, which forces  $\delta_{1i}$  to be estimated identically across all DVs. The same method applies to any additional coefficients for time-varying covariates.

Compare with the univariate individual growth curve model:

$$\text{Level 1 : } Y_{ij} = B_0 + B_1time_{ij} + r_{ij} \quad (4.6)$$

$$\text{Level 2 : } B_0 = \gamma_{00} + \gamma_{01}X_1 + \gamma_{02}X_2 + \dots + \gamma_{0n}X_n + U_0 \quad (4.7)$$

$$B_1 = \gamma_{10} + \gamma_{11}X_1 + \gamma_{12}X_2 + \dots + \gamma_{1n}X_n + U_1$$

where  $Y_{ij}$  is a given performance component for an individual at time  $j$ , and the  $time_{ij}$  term again representing the linear trend of the growth model, and the  $X_n$  terms are between-persons predictors of the slopes and intercepts of the individual growth curves, for example, cognitive ability or conscientiousness, and  $r_{ij}$  and the  $U_j$  terms are residuals at the within- and between-persons levels, respectively (Hofmann, Jacobs, & Baratta, 1993). This model

must be estimated separately for each dependent variable. This univariate treatment does not model the covariance between different performance outcomes. Additionally, this approach does not allow a direct statistical test regarding differences in growth parameters (i.e., H2) or external relationships for the different performance facets (i.e., H4 and H5), but does allow for tests regarding individual differences in performance trajectories (i.e., H3).



## Chapter 5

# RESULTS

### 5.1 Descriptive Statistics

Tables 5.1 and 5.2 present descriptive statistics for the within- and between-person variables, respectively. The sample consisted of 178 police recruits. The mean age of sample respondents was 26.7 ( $sd = 5.5$ ), and predominantly male (91%) and white (88%). Another 5% and 6% reported their race as Black and Latino, respectively. Twenty-eight percent of the sample had previous military experience, most having served as enlisted personnel (90%). Twenty-seven percent of the sample reported some prior law enforcement experience. Fifty-four percent of the sample reported having a bachelor's degree, with 18% and 22% reporting having earned an associate's degree or "some college", respectively.

The performance criteria reported here are Verbal performance (academic performance and report-writing), Physical performance (control tactics and firearms proficiency) and Operations performance (performance in integrated training scenarios, i.e., domestic violence call, night building search, and vehicle stops), contextual performance consisting of dependability, maintenance of discipline and personal appearance, engagement in training, and teamwork and assisting other recruits, examination scores, and instructor ratings of performance on academics, firearms, control tactics and reports (see Section 5.2 for more details on these facets).

Table 5.2 presents the means, standard deviations and intercorrelations for the between person predictors. Observed alpha values are reported in parentheses on the diagonal. Tables 5.3 and 5.4 present the correlations for the level-2 predictors and the performance measures for each time point.

There is no manual for the Big Five Inventory, but large-sample statistics for ages 21-60 are available in Srivastava, John, Gosling, and Potter (2003).

Table 5.1: Descriptive Statistics for level 1 variables for Weeks 3, 6, 9 and 12.

Variable	Week 3		Week 6		Week 9		Week 12	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Extraversion	3.76	0.62	3.73	0.63	3.73	0.64	3.80	0.62
Neuroticism	1.90	0.55	1.88	0.56	1.82	0.57	1.84	0.58
Conscientiousness	4.33	0.49	4.36	0.49	4.37	0.50	4.38	0.51
Agreeableness	4.12	0.51	4.12	0.53	4.13	0.53	4.18	0.52
Openness	3.51	0.56	3.51	0.59	3.55	0.61	3.54	0.60
Social Potency	3.84	0.57	3.78	0.58	3.85	0.57	3.85	0.60
Stress Reactance	1.99	0.67	1.90	0.70	1.88	0.67	1.86	0.67
Self-efficacy	3.32	0.51	3.33	0.48	3.44	0.51	3.46	0.51
Learning GO	4.10	0.55	4.11	0.59	4.14	0.59	4.26	0.52
Prove GO	3.21	0.80	3.14	0.92	2.97	0.96	3.01	0.99
Avoid GO	2.10	0.72	1.95	0.74	1.89	0.75	1.88	0.69
Verbal Perf	3.86	0.64	3.98	0.60	4.07	0.64	4.09	0.64
Physical Perf	3.77	0.69	3.97	0.60	4.08	0.58	4.17	0.57
Ops Perf	3.52	0.85	3.75	0.61	4.02	0.57	4.15	0.54
Contextual Perf	4.59	0.38	4.60	0.40	4.61	0.42	4.59	0.34
Exams	90.27	4.96	91.69	3.89	93.23	3.97	88.39	4.15
Instructor Ratings	8.01	0.49	8.02	0.54	8.01	0.54	8.12	0.64

Table 5.2: Descriptive Statistics for between person variables.

Variable	Mean	Std Dev	Wonderlic	Extrav.	Neurot.	Consc.	Agree	Open
Wonderlic	23.99	5.11	(.88)					
Extraversion	3.76	0.62	-.07	(.81)				
Neuroticism	1.93	0.57	-.04	-.26	(.71)			
Conscientiousness	4.39	0.43	-.07	.18	-.55	(.71)		
Agreeableness	4.19	0.47	-.19	.31	-.45	.49	(.79)	
Openness	3.58	.55	.02	.26	-.18	.18	.26	(.76)

Scale scores from the Srivasatava et al. study for age 27 are: Extraversion (3.28), Neuroticism (3.26), Conscientiousness (3.60), Agreeableness (3.68), and Openness (3.95). The police recruits are slightly higher on Extraversion than is typical for respondents their age, and somewhat lower on Openness. They also report being somewhat more Agreeable, but considerably more Conscientious and less Neurotic.

The findings for Extraversion and Neuroticism are consistent with the expectation that more dominant, less anxious individuals would select to become police officers, though it was also expected that these individuals would be somewhat more impulsive, which is inconsistent with being more conscientious than others of their age. It is possible that the more sensation-

seeking aspect of impulsiveness captured by the above job analysis quotation regarding carrying guns and driving fast may be reflected in the approach-oriented trait of extraversion, while more rule- and norm-consciousness likely present in police officers is being captured by conscientiousness.

Table 5.3: Correlations for major level-2 predictors for weeks 3 and 6

Time	Variable	Ability	Extrav.	Neurot.	Consci.	Agree.	Open.	Verbal	Physical	Ops	Contextual	Exams
0	Ability											
0	Extrav.	-0.04										
0	Neurot.	-0.02	-0.25									
0	Consci.	-0.06	0.2	-0.59								
0	Agree.	-0.2	0.32	-0.49	0.52							
0	Open.	0.07	0.27	-0.21	0.22	0.23						
0	Verbal	0.07	0.18	-0.15	0.12	0.08	0.22					
0	Physical	-0.08	0.24	-0.01	0.06	-0.01	0.21	0.39				
0	Ops	-0.06	0.14	-0.09	0.03	0.06	0.19	0.6	0.48			
0	Contextual	-0.2	0.29	-0.22	0.28	0.32	0.31	0.31	0.32	0.34		
0	Exams	0.36	-0.09	0.05	0.08	-0.09	0.15	0.14	0.07	0.07	-0.08	
0	Ratings	0.12	-0.01	0.03	0.03	-0.14	0	0.05	0.11	0.07	-0.14	0.48
1	Ability											
1	Extrav.	-0.07										
1	Neurot.	-0.02	-0.26									
1	Consci.	-0.04	0.16	-0.55								
1	Agree.	-0.21	0.33	-0.48	0.49							
1	Open.	0.02	0.29	-0.2	0.17	0.25						
1	Verbal	0.18	0.11	-0.16	0.24	0.07	0.12					
1	Physical	-0.02	0.22	-0.06	0.06	0.07	0.12	0.31				
1	Ops	-0.13	0.27	-0.17	0.13	0.11	0.22	0.46	0.57			
1	Contextual	-0.17	0.24	-0.24	0.37	0.27	0.23	0.37	0.23	0.36		
1	Exams	0.31	-0.17	-0.01	0.04	-0.19	-0.02	0.24	-0.09	-0.15	-0.02	
1	Ratings	0.23	-0.12	0.02	0.02	-0.24	-0.03	0.13	0.08	-0.07	-0.02	0.43

Correlations between cognitive ability and Big Five traits with performance measures over time are presented in Tables 5.3 and 5.4. The zero-order correlations showed that cognitive ability began as a poor predictor of all self-rated performance facets, but improved over time for verbal performance. Cognitive ability predicted exam performance at all times, and appeared to increase somewhat as the training program progresses. Cognitive ability also predicted instructor ratings of performance initially, and this prediction appeared to improve over time. Extraversion showed a modest correlation with the initial self-evaluation of verbal performance, which appeared to increase then decrease, but predicted other self-evaluated facets throughout the program; however, Extraversion showed a moderate negative correlation with exam and instructor-rated performance throughout the training. Neuroticism was negatively related to all performance facets at all times, and seemed to be most strongly negatively related to self-evaluations of operations and contextual performance.

Conversely, Conscientiousness displayed positive relationships with all facets, particularly self-evaluations of verbal and contextual performance. Interestingly, Agreeableness seemed to differentiate sources well, positively predicting self-ratings of contextual and operational performance, but showing negative relationships with exam and instructor ratings of performance. Similarly, Openness to Experience predicted self-evaluations of all performance facets, and the Time 0 exam scores, but became slightly negatively related to exam performance over time.

Table 5.4: Correlations with performance for major level-2 predictors for weeks 9 and 12

Time	Variable	Ability	Extrav.	Neurot.	Consci.	Agree.	Open.	Verbal	Physical	Ops	Contextual	Exams
2	Ability											
2	Extrav.	-0.09										
2	Nuerot.	-0.03	-0.26									
2	Consci.	-0.04	0.16	-0.55								
2	Agree.	-0.22	0.33	-0.48	0.49							
2	Open.	0	0.28	-0.2	0.18	0.24						
2	Verbal	0.22	0.08	-0.16	0.26	0.01	0.1					
2	Physical	0.03	0.18	-0.06	0.11	0.03	0.14	0.35				
2	Ops	-0.01	0.18	-0.15	0.14	0.11	0.18	0.5	0.54			
2	Contextual	-0.12	0.23	-0.28	0.4	0.28	0.17	0.35	0.29	0.34		
2	Exams	0.39	-0.11	-0.09	0.1	-0.11	-0.01	0.42	0.06	0.08	0.07	
2	Ratings	0.24	-0.06	0.02	0.01	-0.11	0.01	0.23	0.14	-0.03	0.02	0.44
3	Ability											
3	Extrav.	-0.08										
3	Neurot.	0.02	-0.18									
3	Consci.	-0.05	0.14	-0.56								
3	Agree.	-0.32	0.31	-0.41	0.46							
3	Open.	0	0.27	-0.17	0.21	0.27						
3	Verbal	0.18	0.08	-0.2	0.32	0.11	0.1					
3	Physical	-0.02	0.36	-0.15	0.19	0.1	0.19	0.36				
3	Ops	-0.12	0.29	-0.22	0.23	0.25	0.11	0.48	0.69			
3	Contextual	-0.17	0.16	-0.25	0.41	0.26	0.04	0.28	0.22	0.25		
3	Exams	0.47	-0.11	-0.08	0.17	-0.2	-0.02	0.41	-0.03	-0.03	0.01	
3	Ratings	0.23	0	-0.02	0.21	-0.2	-0.1	0.16	0.07	-0.03	0.09	0.48

## 5.2 Hypothesis testing

### Peformance structure

#### Confirmatory analyses

Hypothesis 1 indicated that the performance indicators would be structured around multiple performance facets. This hypothesis was addressed in sev-

eral ways. First, confirmatory factor analyses were performed on performance data from the first measurement occasion, consisting of 4 self-ratings of contextual performance, self-ratings of academic, firearms, control tactics, and report-writing performance, examination scores, and instructor ratings of academic and firearms performance. CFAs were conducted using the “sem” package (Fox, 2006) in *R* (R Development Core Team, 2008). A hypothesized three independent factors model was compared to a model with only a general factor and to a bifactor model, with a general factor that was independent of three primary factors, consisting of contextual, verbal (academic performance, exams, and report-writing) performance, and physical performance (firearms and control tactics). Results are presented in Table 5.5. The bifactor model provided better fit to the data than did the general or independent factors model, though the fit statistics were still marginal. This bifactor model was then fit to data from each of the subsequent measurement occasions, providing good fit in Weeks 6 and 9, but marginal fit again in week 12. Additionally, there were significant negative variances estimated for week 12, indicating that the model is probably misspecified.

Table 5.5: Confirmatory factor analyses

Model	$\chi^2$	df	CFI	NNFI	RMSEA	SRMR
General Factor	348.97	44	.40	.25	.20	.13
Three Factor	250.69	44	.59	.49	.16	.16
Bifactor (Week 3)	85.74	33	.90	.83	.10	.07
Bifactor (Week 6)	71.33	33	.92	.86	.08	.07
Bifactor (Week 9)	61.68	33	.96	.93	.07	.07
Bifactor (Week 12)	79.31	33	.91	.86	.09	.07

### Item cluster analyses

Since the multivariate multilevel model uses observed variables, Revelle’s (1979) item clustering approach was applied within each measurement occasion separately and to all occasions simultaneously. The intent behind this procedure was to construct scales in a way that was consistent with items’ observed variability, as opposed to the latent structure approach used in the confirmatory factor models. Figure 5.1 through 5.3 present the dendrograms for the item cluster analyses. Each cluster ellipse in these figures presents

the cluster's estimated Cronbach's (1951) alpha and Revelle's (1979) beta, an estimate of general factor saturation for the cluster. In all cluster analysis results that follow, if an item's name is followed by a minus sign, that item displays a negative loading on the general factor.

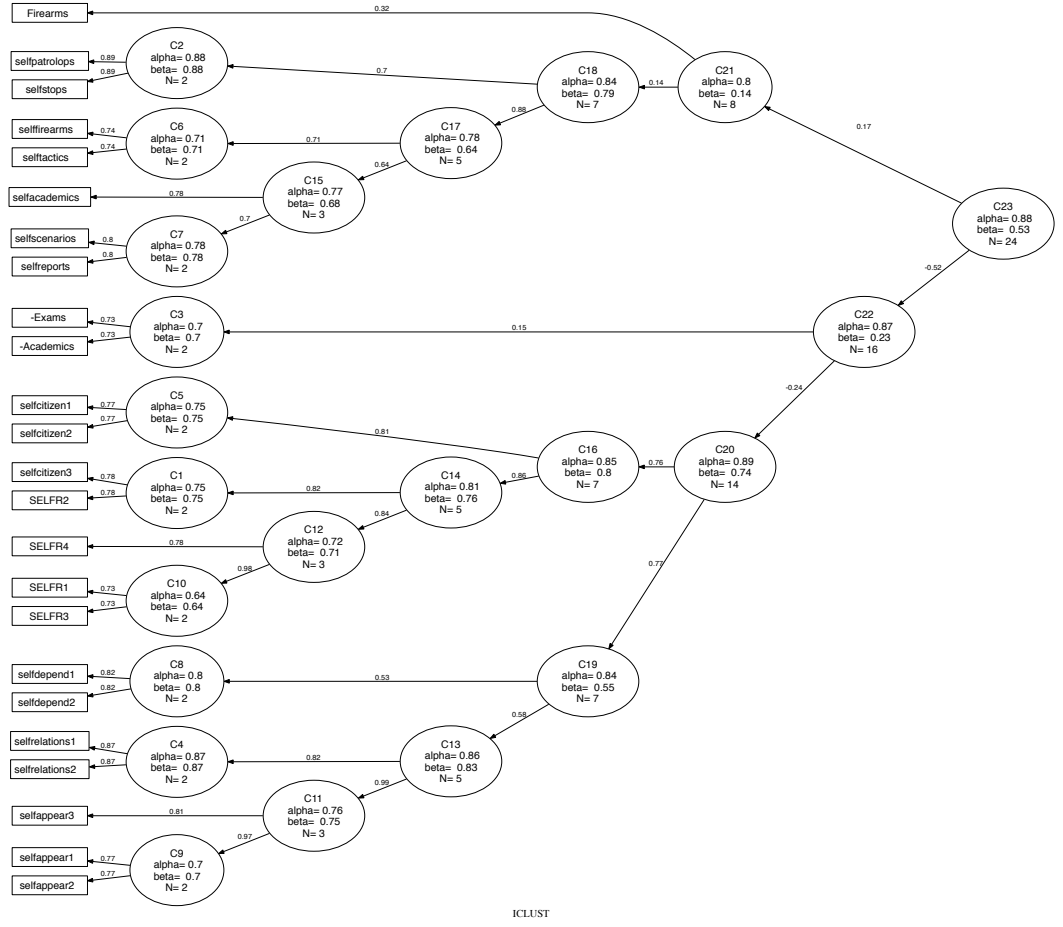
There are two points to note about the item cluster analyses. The first is that the clustering over all measurements occasions produces three major clusters: self-rated task performance, self-rated contextual performance, and instructor-rated task performance with examination scores. Note also that instructors did not rate contextual performance variables. The other feature to notice is that performance seems more undifferentiated at week 3 (Time 0) than at other measurement occasions, specifically, there is a single cluster at the root of the dendrogram for week 3, but later times are unrooted. Looking at which items cluster together, it appears that the two major clusters at later measurement occasions represent task performance (e.g., exams, academics, firearms) and contextual performance (e.g., citizenship and maintaining proper appearance). The cluster pattern for week 12 is very similar to that for the clustering over all measurement occasions, which appears to differentiate task from contextual performance, but also self- from instructor-rated task performance. Examining Figure 5.2, the task performance cluster has two main sub-clusters, self-ratings, and instructor-ratings combined with examination scores. The self-ratings themselves show three main sub-clusters, with academic performance and report writing grouping together, firearms and control tactics grouping together, and patrol operations and vehicle stops linking quickly then joined by integrated scenarios. Additionally, the "other"-source sub-cluster shows similar patterning, in that exams and instructor-rated academic performance link first, and then link with instructor-rated firearms performance.

## Time trends

### Exploring the time trends of performance indicators

The temporal patterning of these performance indicators was addressed using univariate multilevel modeling (Bliese & Ployhart, 2002). This approach models the effect of time by regressing the dependent variable onto the linear

Figure 5.1: Item cluster analysis for Week 3



index of time (coded as 0, 1, 2, 3):

$$\text{Level 1 : } Y_{ij} = B_0 + B_1 \text{time}_{ij} + r_{ij} \quad (5.1)$$

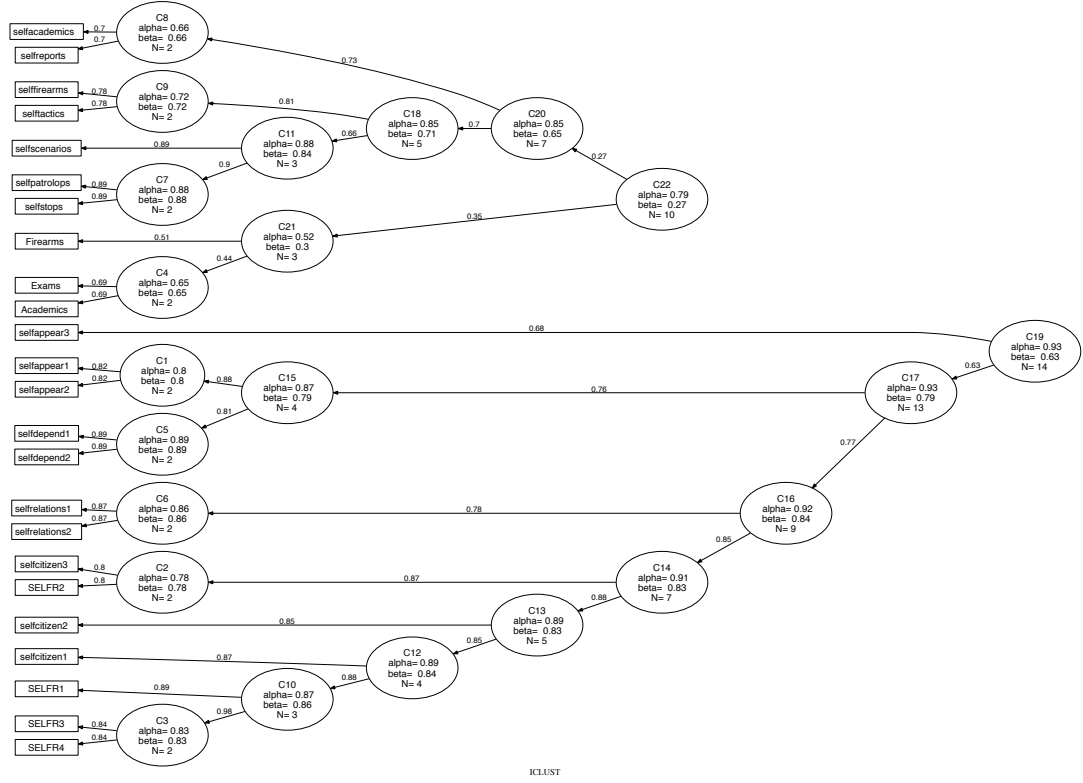
$$\text{Level 2 : } B_0 = \gamma_{00} + U_0 \quad (5.2)$$

$$B_1 = \gamma_{10} + U_1$$

Random effects are modeled as variance in the  $U$  terms in Equation 5.2.  $U_0$  allows for individual differences in intercepts, and  $U_1$  allows for individual differences in the slope on time.

The results are presented in Table 5.6. This table presents unstandardized

Figure 5.2: Item cluster analysis for Week 9



coefficients. “Trend” refers to the fixed effect for a linear time trend (i.e.,  $\gamma_{10}$ ), coded as 0, 1, 2, 3, and “Random” refers to the random effect around the trend. The fixed effect is essentially the mean trend over time, and the random effect reflects individual variability about the trend.

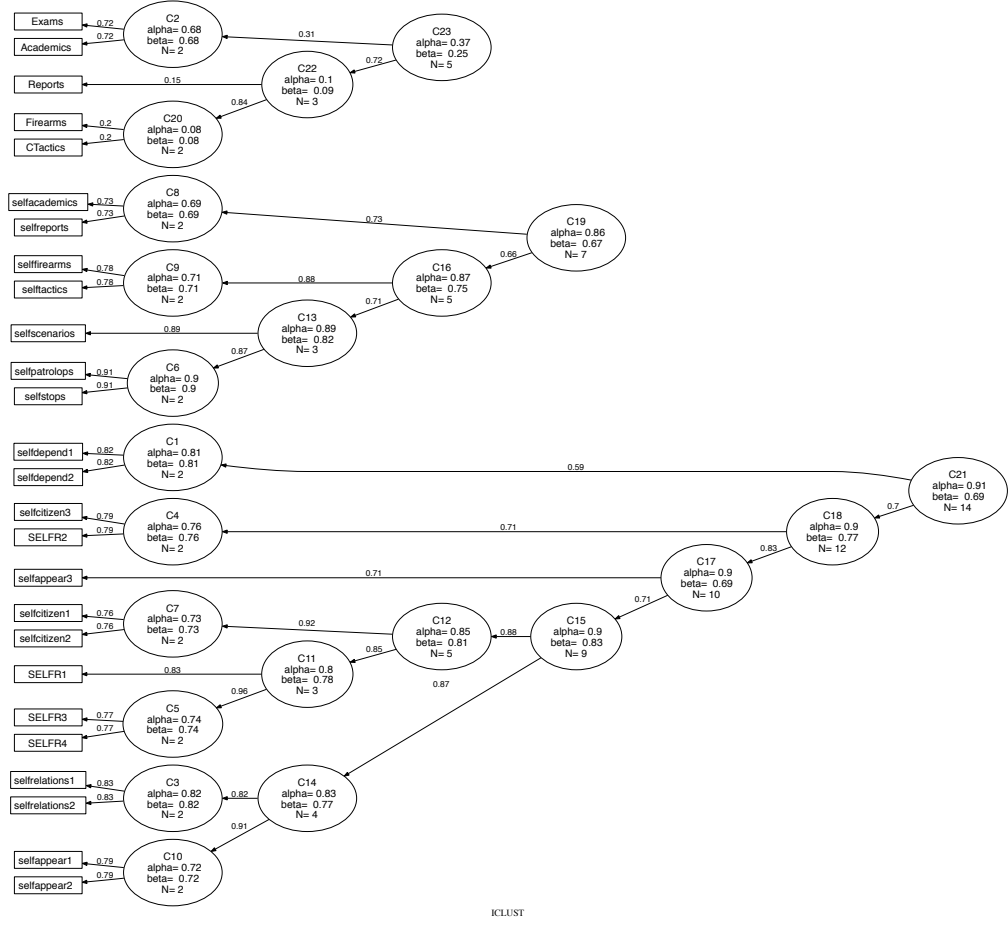
The likelihood ratio test for the trend compares a model with a fixed effect for time to a model with only a random intercept, such that:

$$\begin{aligned}
 \text{Level1 : } Y_{ij} &= B_0 + B_1 \text{time}_{ij} + r_{ij} \\
 \text{Level2 : } B_0 &= \gamma_{00} + U_0 \\
 B_1 &= \gamma_{10}
 \end{aligned} \tag{5.3}$$

for the full model is compared to:



Figure 5.3: Item cluster analysis for all times



$$\text{Level1 : } Y_{ij} = B_0 + r_{ij} \quad (5.4)$$

$$\text{Level2 : } B_0 = \gamma_{00} + U_0$$

for the reduced model.

The likelihood ratio test for the random effect compares a model with a

random effect for time to one with only a fixed effect for time, such that:

$$\begin{aligned}
\text{Level1 : } Y_{ij} &= B_0 + B_1 \text{time}_{ij} + r_{ij} \\
\text{Level2 : } B_0 &= \gamma_{00} + U_0 \\
B_1 &= \gamma_{10}
\end{aligned} \tag{5.5}$$

is compared to a model with  $B_1 = \gamma_{10} + U_1$  (i.e., the model presented in Equations 5.1 and 5.2).

All of the fixed effects for linear time trends were significant (i.e., significant parameter estimate for  $\gamma_{10}$ ), except for instructor ratings of report writing. All of the random effects (i.e., variance of  $U_1$ ) were significant except for exam scores and instructor-rated academic performance. Self-ratings of academic performance and report-writing showed very small positive trends, and self-rated firearms and control tactics performance showed slightly larger trends. Exams showed a small negative trend for time. All of the time trends and random effects for instructor ratings were very small in magnitude, reflecting the limited variance in these indicators (see Table 5.1).

#### Time trends of cluster-analytic scales

Finally, the slopes for the performance facets based on the cluster analysis results were explored, in the same manner described in the previous section.

Tables 5.7 and 5.8 present the results of these basic multilevel models. These are examined at the sub-cluster level, using self-evaluations of verbal, physical, operational, and contextual performance, as described above, instructor-ratings, and examination scores.

The intraclass correlation (ICC) for the null model estimates the amount of total variance in responses attributable to differences between individuals, with  $1 - \text{ICC}$  indicating the percentage of variance within-person. Hence, a low ICC indicates that most of the variability is within-person, whereas a high ICC indicates that most of the variability is between individuals. The “null model” is one with only a random intercept for individuals, such that:

Table 5.6: Univariate change over time in performance indicators

Dependent Variable	Model	Likelihood Ratio	$p$	coefficient	$p$
Exams	Trend	11.63	.001	-.410 <sup>a</sup>	.001
	Random	< .0001	1.000		
Self-rated Academics	Trend	15.39	.000	.070 <sup>a</sup>	.001
	Random	15.14	.001	.149 <sup>b</sup>	.001
Self-rated Firearms	Trend	38.72	.000	.115 <sup>a</sup>	.000
	Random	14.28	.001	.169 <sup>b</sup>	.001
Self-rated Control Tactics	Trend	73.44	.000	.144 <sup>a</sup>	.000
	Random	9.50	.009	.136 <sup>b</sup>	.009
Self-rated Reports	Trend	17.38	.000	.080 <sup>a</sup>	.000
	Random	10.49	.005	.147 <sup>b</sup>	.005
Instr-rated Academics	Trend	7.40	.065	.027 <sup>a</sup>	.007
	Random	0.03	.980		
Instructor-rated Firearms	Trend	21.37	.000	-.029 <sup>a</sup>	.000
	Random	67.48	.000	.072 <sup>b</sup>	.000
Instr-rated Control Tactics	Trend	11.91	.001	.016 <sup>a</sup>	.000
	Random	246.42	.000	.065 <sup>b</sup>	.000
Instr-rated Reports	Trend	0.30	.586	-.003 <sup>a</sup>	.705
	Random	188.66	.000	.067 <sup>b</sup>	.000

<sup>a</sup> This is the estimate of  $\gamma_{10}$ .

<sup>b</sup> This is the estimate of the variance of  $U_1$ .

$$\text{Level1 : } Y_{ij} = B_0 + r_{ij} \quad (5.6)$$

$$\text{Level2 : } B_0 = \gamma_{00} + U_0$$

so that the estimate of the variance in  $r_{ij}$  represents the within-person variance and the variance of  $U_0$  represents the between-person variance, allowing for calculation of the ICC.

The likelihood ratio tests presented in these tables compare a model with a random intercept and random slope for time to a model with a random intercept, but only a fixed effect for time, so that that model comparison is the same as that outlined for Equation 5.5 above. The fixed effect estimate for the coefficient of time ( $\gamma^*\text{time}$ ; i.e., the linear growth parameter) is also presented.

Table 5.7: ICCs, Likelihood Ratio tests, and fixed effect estimates of performance measures

Dependent Variable	ICC	Model	LogLike	LR	$p$	$\gamma^*\text{Time}$
Verbal Perf	.55	Rand Int	-504.87			
		Rand Int,Slp	-493.92	21.90	.00	0.08
Physical Perf	.50	Rand Int	-492.96			
		Rand Int,Slp	-485.14	15.65	.00	0.13
Ops Perf	.32	Rand Int	-585.31			
		Rand Int,Slp	-565.21	40.20	.00	0.22
Contextual Perf	.61	Rand Int	-163.61			
		Rand Int,Slp	-155.46	16.30	.00	0.00
Exams	.39	Rand Int	-2030.25			
		Rand Int,Slp	-2030.25	0.00	1.00	-0.41
Instructor Ratings	.59	Rand Int	-448.33			
		Rand Int,Slp	-433.18	30.30	.00	0.085

All of the performance domains show substantial within-person variability (at least 39% of the variability is within-person). As can be seen in the column  $\gamma^*\text{Time}$  in Table 5.7, there are small positive slopes for verbal and physical performance, and a fairly substantial positive slope for operational performance, but no average change for contextual performance. These effects will be explored more fully in the multivariate tests section below. There is also substantial negative change in exam scores and no evidence for indi-

vidual variability about the fixed effect.

The instructor ratings are more problematic. First, there is little variance in instructor ratings, period (Table 5.1). Instructors rated recruits qualitatively, from “remedial” performance in a domain, such as firearms or control tactics, to “excelling”, with most recruits typically receiving ratings of “making adequate progress”. These ratings were scored with remedial equal to 1, making progress equal to 2, and excelling equal to 3. These ratings are explained more in Section 5.2 below.

### Time trends of predictors

Table 5.8 is set up identically to table 5.7. The ICCs indicate substantial within-person variability for all of the within-person covariates, except the Big-5 and Iowa personality scales, which show substantial coherence over measurements. Small negative trends were identified for neuroticism, stress reactance, performance-prove and avoidance goal orientations. Small positive trends were found for agreeableness, learning goal orientation, and self-efficacy. Individual variability around these fixed growth parameters was identified for all variables except stress reactance and learning goal orientation.

Note that the directions of observed changes were generally in accordance with the ancillary predictions. Neuroticism and stress reactance showed decreases, while Social Potency showed an increase. Also, the slopes were larger for narrower measures, such as Social Potency and Stress Reactance, than for broader measures, such as Extraversion and Neuroticism. There was no random effect on Stress Reactance. This suggested that changes in Stress Reactance, while small, were essentially the same over recruits, which is consistent with a socialization process as described in Section 2.8.1.

These tests suggested some small systematic change in personal characteristics and performance during the training program. They also indicated large within-person variability in self-perceptions and performance. Given these justifications, the first set of tests examined the growth curves. Then the unsystematic within-person variability was be modeled.

Table 5.8: ICCs, Likelihood Ratio tests, and fixed effect estimates of time-varying predictor change

Dependent Variable	ICC	Model	LogLike	LR	$p$	$\gamma^*\text{Time}$	$p(\gamma)$
Extraversion	0.86	Rand Int	-260.08				
		Rand Int,Slp	-251.19	17.80	.00	0.01	.15
Neuroticism	0.80	Rand Int	-258.79				
		Rand Int,Slp	-243.55	30.48	.00	-0.03	.01
Conscientiousness	0.76	Rand Int	-220.06				
		Rand Int,Slp	-193.06	53.99	.00	0.02	.15
Agreeableness	0.83	Rand Int	-167.92				
		Rand Int,Slp	-158.72	18.40	.00	0.02	.01
Openness	0.86	Rand Int	-210.15				
		Rand Int,Slp	-198.67	22.96	.00	0.01	.10
Social Potency	0.78	Rand Int	-308.86				
		Rand Int,Slp	-304.87	7.99	.02	0.07	.20
Stress Reactance	0.71	Rand Int	-470.13				
		Rand Int,Slp	-469.92	0.42	.81	-0.04	.00
Learning GO	0.60	Rand Int	-411.97				
		Rand Int,Slp	-411.98	0.03	.99	0.05	.00
Prove GO	0.64	Rand Int	-720.63				
		Rand Int,Slp	-713.23	14.81	.00	-0.08	.00
Avoid GO	0.53	Rand Int	-621.40				
		Rand Int,Slp	-618.27	6.26	.04	-0.08	.00
Self-efficacy	0.59	Rand Int	-333.98				
		Rand Int,Slp	-330.45	7.04	.03	0.05	.00

## Multivariate multilevel modeling

The major hypotheses of the study were examined by fitting the multivariate multilevel model as described in 4.1 and 4.2. The multivariate tests were restricted to the self-reported performance facets. First, these displayed far greater variability than the instructor ratings. Second, the recruits and instructors provided ratings on the same scales during the final week of training, so the agreement between self and instructor could be estimated (see Table 5.16 below). Two items had to be dropped due to zero variance in the instructor ratings, but the average between-source correlation was .11 (median .11,  $sd = .11$ ), with a maximum correlation of .35 and a minimum of -.06.

Hypotheses 2 and 3 regarded differences in growth curves, specifically that

the growth curves would differ over performance facets (H2) and that there would be individual differences in the parameters of the growth curves. To test this hypothesis, SAS PROC MIXED was used to obtain maximum likelihood estimates of the model displayed in Equations 4.1, 4.2, and 4.3. The results were used to perform likelihood ratio tests for two nested models: (a) a model with the one intercept and one slope for all performance facets (i.e., all performance facets show identical change patterns), and (b) a model with a unique intercept for each facet, but with the slopes constrained to equality, as presented in Equation 4.5. Both are compared to the hypothesized model, which allows unique fixed effects on time for each performance facet (i.e., the regression coefficient for the linear time index is freely estimated for each dependent variable; see Equation 4.2). The  $-2 \times \log$ -likelihood for the hypothesized model was 2943.3, for model (a) it was 4181.5, and for model (b) it was 3163.1.

The likelihood ratio comparing the hypothesized model to model (a) was 1238.2, which is distributed as a  $\chi^2$  with 4 degrees of freedom ( $p < .001$ ). The likelihood ratio test comparing the hypothesized model to model (b) was 219.8, which is distributed as a  $\chi^2$  with 2 degrees of freedom ( $p < .001$ ). These tests indicate that removing both unique random intercepts and slopes results in a significant reduction in explanation for outcomes. There is also substantial reduction in explanation when the fixed effects of the growth curves' slopes are constrained to be equal. This indicates full support for hypothesis 2. These results mean that there are different intercepts for self-ratings of verbal, physical, operational, and contextual performance. The differences in slopes indicate that these performance facets change at different rates.

In order to more fully explore what these slopes mean, the hypothesized model was refit using restricted maximum likelihood estimation (REML). The fixed effects estimates are presented in Table 5.9. The slope for operational performance has the largest magnitude, followed by physical, followed by verbal. Verbal is probably the most similar to performance in academic settings, as it is focused on exams and written reports. Physical is more focused on aspects of police work, such as control tactics, and operational performance focuses on performance that is almost exclusively oriented toward police work. This pattern of results is consistent with the context of a police training program, such that the part of performance that is most

unique to the context shows the most development.

The fixed effect for contextual performance is negligible, indicating no systematic change for participants over the course of the study. This is to be expected as the police training program is focused on learning various police tasks, with relatively little emphasis on citizenship and contextual performance behaviors. This leaves the recruits more or less to their own devices regarding the development of contextual performance.

Table 5.9: REML estimates of fixed effects for unique effects of time

Effect	Fixed Effects				
	Estimate	Std Error	DF	t-value	<i>p</i>
Verbal Intercept	3.89	0.032	2467	121.3	< .0001
Physical Intercept	3.81	0.044	2467	86.5	< .0001
Ops Intercept	3.54	0.048	2467	73.6	< .0001
Contextual Int	4.59	0.029	2467	159.86	< .0001
Verbal Slope	0.07	0.024	2467	3.15	.0016
Physical Slope	0.13	0.017	2467	7.47	< .0001
Ops Slope	0.21	0.022	2467	9.54	< .0001
Contextual Slp	-0.0001	0.010	2467	-0.10	.9230

The question of individual differences in intraindividual change, such as that posed by hypothesis 3 and the above speculations regarding contextual performance, can be addressed by examining the random effects. For instance, a significant variance estimate for the slope of contextual performance on time supports the possibility that there are meaningful individual differences in the pattern of change in contextual performance. Even small effects may signify important differences in individual trajectories that may be predicted using between-person predictors (e.g., cognitive ability or personality measures).

Table 5.10 presents the restricted maximum likelihood (REML) estimates for the random effects for the multivariate growth model. The covariance parameter estimates for all intercepts, slopes, and residuals were significant. The covariance between random verbal intercepts and verbal slopes was not significant, which suggests that where an individual starts on his or her verbal performance did not restrict change.

The effects for contextual performance's slope were all negligibly small. The only large random effects appeared to occur for the intercepts of verbal,



physical, and operations performance and their covariances. These covariances were all positive. Recall that time is coded as 0, 1, 2, and 3, so the intercepts reflect performance at the first measurement occasion, so these positive covariances probably reflect greater influence of a general factor during the first measurement occasion (Figure 5.1). Variability in slopes and intercepts supported hypothesis 3, that there are individual differences in performance trajectories.

There was little covariance among the residual terms. Covariance among the residuals would suggest that the level for one performance facet is related to the the level for another *within* person. These parameters answer questions about how two variables move together for individuals, which addresses the assumption of whether performance behaviors are exchangeable. The lack of substantial covariance here suggested that performances in the different domains were largely independent of one another, even when measured at the same time. This is actually inconsistent with the argument above that performance behaviors are non-fungible. We would expect negative covariances between facets of performance if that were the case. For example, when an individual was assisting a coworker, he or she could not also be doing his or her own work; therefore a positive peak on contextual performance would result in a negative peak for task performance. This does not appear to be the case for these data. Note though, that these were self-evaluations of performance, not actual performance. Further, these measurements were only taken every three weeks, not as an “on-line” assessment, so such relationships may be obscured.

Hypothesis 4 stated that the influence of regulatory and identity mechanisms would differ for the different performance facets. The models tested here refit the growth models allowing the within-person covariates to explain residual within-person variance, controlling for that trend. The full models with these covariates for each facet were compared to nested models with the slopes for the covariates restricted to equality for each facet (i.e., the influence of the mechanism is the same for all four facets of performance).

The likelihood ratio tests for Work Goal Orientation, Self-efficacy, Social Potency, and Stress Reactance, and the Big Five Personality Traits are reported in Table 5.11. These time-varying covariates were tested using the procedures outlined in Chapter 4. The models fitted were

$$Y_{tid} = \sum_{d=1}^D DV_d(\delta_{0id} + \delta_{1id}time_{ti} + \delta_{2id}X_{tid} + e_{tid}) \quad (5.7)$$

for the full model, which was compared to

$$Y_{tid} = \sum_{d=1}^D DV_d(\delta_{0id} + \delta_{1id}time_{ti} + e_{tid}) + \delta_{2i}X_{ti} \quad (5.8)$$

for the reduced model, as in Equation 4.5.

The influence of the covariates were modeled as fixed effects. Additionally, the full models for Learning Goal Orientation and Stress Reactance (noted with asterisks in Table 5.11) did not produce positive definite covariance matrices for the between-persons random effects. Therefore, the estimates for the full model should not be interpreted. The regression coefficient on learning goal orientation for the reduced model, one where the effect is uniform across performance facets, was 0.22,  $p < .0001$ . The coefficient for stress reactance was -0.09,  $p = .01$ . Table 5.12 presents the restricted maximum likelihood estimates for the regression coefficients for the time-varying covariates.

All of the tests presented in 5.11 were statistically significant. In all cases, the reduced model increased the  $-2 \times (\log\text{-likelihood})$ , indicating inferior fit to the full model. This suggested that the influences of the regulatory and identity mechanisms tested differed over the performance facets, supporting hypothesis 4. In addition, the pattern of results seemed consistent with information from cross-sectional validation work. Negative characteristics, such as Neuroticism and Avoidance goal orientation showed negative associations with outcomes, particularly contextual performance. Further, Neuroticism was negatively associated with verbal and operational performance facets. Performance-prove goal orientation was positively associated with operational performance only, whereas conscientiousness was positively associated with verbal, physical, and contextual performance. As expected, Social Potency was associated with all performance facets, as was its associated Big-5 domain, Extraversion<sup>2</sup> and Self-efficacy. Agreeableness was only associated with contextual performance, and Openness was associated with all aspects of performance, likely due to the fact that this was a training

---

<sup>2</sup>The REML coefficients in Table 5.12 were not estimated simultaneously, so one should not compare their magnitudes across rows.

context.

Hypothesis 5 was tested in a very similar manner, with time-invariant covariates predicting the intercepts and slopes of the individual performance facets' trajectories in the full models compared to two reduced models. "No slopes" models constrained the growth parameters ( $B_{1s}$ ) to equality as above. "No int or slope" models constrained both the intercepts and slopes ( $B_{0ids}$  and  $B_{1ids}$ ) to equality over dependent variables. Likelihood ratio tests were conducted for cognitive ability and Big-5 personality traits. Results are presented in Table 5.13. The likelihood ratios are distributed as  $\chi^2$  with 3 degrees of freedom for the "no slopes" tests, and 6 degrees of freedom for the "no int or slp" tests. Table 5.14 presents the fixed effect estimates of the time invariant covariates.

All of the likelihood ratio tests comparing the full models to the "no int or slp" models were significant, indicating that removing both the differentiated intercepts and slopes between performance facets significantly reduces the explained variance. None of the likelihood ratio tests comparing the full models to the "no slopes" models were significant. This indicated that the effect of these between-person predictors was uniform over performance facets' slopes, but unique for their intercepts. The coefficients presented in Table 5.14 should be interpreted as regression coefficients explaining variability in the random effects. The slope column should be interpreted in much the same way a fixed effect is, as the average effect of the predictor for all performance facets' slopes (i.e., it is a fixed effect for level 2 in Equations 4.1 through 4.3).

Interestingly, cognitive ability was unrelated to most of the model parameters, emerging as only a significant, and negative, predictor of the contextual performance intercept ( $\gamma = -0.016, p = .002$ ). Recall, however, that the zero-order correlations for cognitive ability and the self-evaluations of performance were negligible. Extraversion, on the other hand, significantly predicted all facets' intercepts but not the general slope. Similarly, neuroticism did not predict the facets' slope, but did negatively predict the intercepts for verbal, operational, and contextual performance. Conscientiousness predicted the intercept for verbal and contextual performance, and neared conventional significance levels for predicting the general performance slope. Agreeableness only predicted the intercept of contextual performance. Openness emerged as a predictor of all facets' intercepts, and as the only

conventionally significant negative predictor of the general performance trajectory.

### Further univariate tests

The item cluster analyses suggested that instructor ratings should probably not be treated as a multivariate criterion. A unit-weighted composite was constructed. The univariate tests will treat the instructor ratings and exam separately from the above. The exams are on a vastly different scale than all rated measurements. Since the design allows treating time one as a relatively pure intercept, coded 0, the judgment was made not to center the dependent variables. Doing so would require interpreting all effects as deviations from mean performance, which was deemed problematic for a true intercept model. Univariate growth models were fit using the *R* “multilevel” (Bliese, 2002) and “nlme” (Pinheiro & Bates, 2000) packages.

Sixty-two percent of the total variance in exam scores was within person (null model ICC = .38). A growth model with a random intercept was fit to the data (as presented in Equation 5.5 above), which produced a random intercept estimate of 8.36 ( $p < .0001$ ), and a regression coefficient for time of -0.41 ( $p = .007$ ). When a random slope model was tested, it produced a random effect estimate equal to 0 for time, meaning that there was effectively no variability in the level-2 residual term for time ( $U_1$  in Equation 4.7). Between-person predictors were used in an attempt to explain exam score intercept variability. For each of these the model with a random intercept and a fixed slope on time but no covariate serves as the basis for comparison (i.e., Equation 4.6, with the between person predictors entered only for  $B_0$  and no  $U_1$  term).

Predicting the intercept of exams with cognitive ability produced a Likelihood ratio of 49.015, with 2 degrees of freedom ( $p < .0001$ ;  $\gamma = 0.35, p < .0001$ ). Extraversion produced a LR of 4.31, with 1 degree of freedom ( $p = .04$ ;  $\gamma = -0.90, p = .04$ ). Neuroticism did not predict exam intercepts, nor did conscientiousness or openness, but Agreeableness did (LR=5.14, df=1,  $p = .02$ ;  $\gamma = -1.25, p = .02$ ). Most time-varying covariates did not predict concurrent exam scores. Performance-prove goal orientation did (LR=3.72,  $p = .05$ ;  $\gamma = -0.47, p = .05$ ).

Next, instructor ratings of performance was examined. There was no variance in ratings of performance in vehicle stops or patrol operations simulations, or scenario-based training, and limited variability in the remaining ratings of academics, firearms, control tactics and reports. A sum composite of these ratings was formed to preserve some variability. The null model produced an ICC estimate of .59, suggesting substantial within-person variability. The likelihood ratio test for adding a linear time trend was non-significant (see Table 5.7).

Within-person regulatory mechanisms were used in an attempt to explain some of the within-person variability in instructor ratings (i.e., the level-1 model was  $Y_{ij} = B_0 + B_1X_{ij} + r_{ij}$ , where  $X_{ij}$  is the  $i^{th}$  participant's standing on the regulatory mechanism, X, at time  $j$ ). Performance-prove goal orientation predicted instructor ratings (LR=12.20, 1 df,  $p < .0001$ ), and further tests supported a random slope on this orientation (LR=306.96, 1 df,  $p < .001$ ), with the fixed effect = -.04 and a variance of .50. Self-efficacy was also supported as a time-varying covariate (LR=5.44, 1 df,  $p = .02$ ) with a random slope on efficacy (LR=294.66, 1 df  $p < .0001$ ), with the fixed effect = -.09 and variance of .64.

Cognitive ability emerged as a level-2 predictor of the intercept for instructor ratings, this means that it predicted initial instructor ratings on the composite of academics, firearms, control tactics, and report-writing (LR=9.27, 1 df,  $p = .002$ ), though its coefficient was small (.02). Between-persons Agreeableness again predicted instructor ratings' intercepts (LR=7.39, 1 df,  $p = .007$ ),  $\gamma = -0.21$ .

### 5.3 Slopes as predictors

As an initial investigation of whether an individual's trajectory on knowledge or skill matters for latter outcomes, the correlation of random intercepts and slopes with several "level-2" (i.e., between-person) outcomes was examined. These results report the relationships between estimates of individual's unique slopes and intercepts with the outcomes. Intercepts and slopes for exam scores, as an index of knowledge, and self- and instructor-rated performance in simulation exercises, as an index of skill were used as predictors of total hours absent during the training program, passing or failing the state

certification exam for police officers, scores on the firearms written exam, the score on the practical firearms exam, the training program firing range rating, awards received, and final overall instructor ratings. Results are presented in Table 5.15.

Examination of the table shows that the intercept for exams during the training program predicted later examination scores and final instructor evaluations very well (.34, .36, and .49, respectively), but that the slopes on time did not predict other outcomes well, except for final instructor ratings, which it predicts negatively. This is likely a ceiling effect, i.e., people who start off well have no place to go, whereas people who start off poorly have room to improve.<sup>5</sup> Self skill ratings have small associations with firing range scores and final instructor ratings, and their slopes show very small negative associations with absentee hours. Instructor skill ratings and slopes show exactly the same pattern of results, perhaps due to the low variance in instructor ratings, and predict firearms exams, firing range scores, awards, and, unsurprisingly, final instructor ratings well. Note that instructor ratings and intercepts are essentially perfectly correlated, so knowledge of the slope provides no additional information regarding later outcomes. Therefore, the slopes of these performance determinants seem to provide little information regarding later outcomes.<sup>6</sup>

## 5.4 Self-Instructor Correlations.

The associations between self- and instructor-ratings during week 12 are presented (Table 5.16). Self ratings are listed vertically and instructor ratings are listed horizontally in the following table.

---

<sup>5</sup>This is similar to the standard finding in growth curve analyses of, say, GPAs that intercepts and slopes have negative correlations (e.g., Zyphur et al., 2008).

<sup>6</sup>Field training officer ratings of on-the-job performance are also being collected, but these data will not be available for some time.

Table 5.10: REML estimates of random effects for individual differences in growth curves

Parameter	Subject	Random Effects			
		Variance Estimate	Error of Estimate	z-Value	<i>p</i>
Verbal Intercept	Between	0.21	0.03	6.05	<.0001
Cov(Phys Int,Verbal Int)	Between	0.11	0.03	3.96	<.0001
Physical Intercept	Between	0.26	0.04	6.48	<.0001
Cov(Ops Int,Verbal Int)	Between	0.20	0.04	5.49	<.0001
Cov(Ops Int,Phys Int)	Between	0.19	0.04	4.91	<.0001
Operations Intercept	Between	0.37	0.06	6.59	<.0001
Cov(Ctxtl Int,Verbal Int)	Between	0.05	0.02	3	0.0027
Cov(Ctxtl Int,Phys Int)	Between	0.06	0.02	3.35	0.0008
Cov(Ctxtl Int,Ops Int)	Between	0.09	0.02	3.86	0.0001
Contextual Intercept	Between	0.12	0.02	7.01	<.0001
Cov(Verb Slp,Verbal Int)	Between	-0.01	0.01	-0.69	0.4909
Cov(Verb Slp,Phys Int)	Between	-0.03	0.01	-2.35	0.0186
Cov(Verb Slp,Ops Int)	Between	-0.04	0.01	-3.26	0.0011
Cov(Verb Slp,Ctxtl Int)	Between	-0.01	0.01	-0.76	0.4457
Verbal Slope	Between	0.02	0.01	3.43	0.0003
Cov(Phys Slp,Verbal Int)	Between	-0.01	0.01	-1.22	0.2223
Cov(Phys Slp,Phys Int)	Between	-0.03	0.01	-2.65	0.008
Cov(Phys Slp,Ops Int)	Between	-0.02	0.01	-1.85	0.0641
Cov(Phys Slp,Ctxtl Int)	Between	0.00	0.01	-0.29	0.7696
Cov(Phys Slp,Verb Slp)	Between	0.01	0.00	3.08	0.0021
Physical Slope	Between	0.02	0.01	3.41	0.0003
Cov(Ops Slp,Verbal Int)	Between	-0.03	0.01	-2.22	0.0261
Cov(Ops Slp,Phys Int)	Between	-0.03	0.01	-1.84	0.0658
Cov(Ops Slp,Ops Int)	Between	-0.09	0.02	-4.65	<.0001
Cov(Ops Slp,Ctxtl Int)	Between	-0.01	0.01	-1.13	0.2567
Cov(Ops Slp,Verb Slp)	Between	0.02	0.01	3.23	0.0013
Cov(Ops Slp,Phys Slp)	Between	0.02	0.01	2.88	0.004
Operations Slope	Between	0.04	0.01	4.6	<.0001
Cov(Ctxtl Slp,Verbal Int)	Between	0.00	0.01	0.25	0.802
Cov(Ctxtl Slp,Phys Int)	Between	-0.01	0.01	-0.94	0.3469
Cov(Ctxtl Slp,Ops Int)	Between	-0.02	0.01	-2.12	0.0344
Cov(Ctxtl Slp,Ctxtl Int)	Between	-0.01	0.00	-2.71	0.0067
Cov(Ctxtl Slp,Verb Slp)	Between	0.00	0.00	1.26	0.2066
Cov(Ctxtl Slp,Phys Slp)	Between	0.00	0.00	0.39	0.6961
Cov(Ctxtl Slp,Ops Slp)	Between	0.00	0.00	1.43	0.1524
Contextual Slope	Between	0.01	0.00	3.45	0.0003
Verbal Residual	Within	0.14	0.01	12.79	<.0001
Cov(Phys Res,Verb Res)	Within	0.03	0.01	4.05	<.0001
Physical Residual	Within	0.14	0.01	12.75	<.0001
Cov(Ops Res,Verb Res)	Within	0.06	0.01	5.94	<.0001
Cov(Ops Res,Phys Res)	Within	0.05	0.01	5.13	<.0001
Operations Residual	Within	0.18	0.01	12.75	<.0001
Cov(Ctxtl Res,Verb Res)	Within	0.02	0.00	4.96	<.0001
Cov(Ctxtl Res,Phys Res)	Within	0.01	0.00	2.02	0.0437
Cov(Ctxtl Res,Ops Res)	Within	0.02	0.01	3.42	0.0006
Contextual Residual	Within	0.05	0.00	12.93	<.0001

Table 5.11: Likelihood ratio tests for unique effects of within-person covariates.

Model	$-2 \times \text{Log-likelihood}$	$\Delta\chi^2$	$p$
<hr/> Self-efficacy <hr/>			
full	2824.7		
reduced	3055.1	230.4	.000
<hr/> Learning GO* <hr/>			
full*	2903.3		
reduced	3102.5	199.2	.000
<hr/> Prove GO <hr/>			
full	2933.8		
reduced	3181.6	247.8	.000
<hr/> Avoid GO <hr/>			
full	2934		
reduced	3169.6	235.6	.000
<hr/> Social Potency <hr/>			
full	2918.1		
reduced	3142.3	224.2	.000
<hr/> Stress Reactance* <hr/>			
full	2934.6		
reduced	3177.3	242.7	.000
<hr/> Extraversion <hr/>			
full	2922.8		
reduced	3148.2	225.4	.000
<hr/> Neuroticism <hr/>			
full	2925.7		
reduced	3158.6	232.9	.000
<hr/> Conscientiousness <hr/>			
full	2906.2		
reduced	3121.4	215.2	.000
<hr/> Agreeableness <hr/>			
full	2923.5		
reduced	3146.2	222.7	.000
<hr/> Openness <hr/>			
full	2931.7		
reduced	3157.6	225.9	.000



Table 5.12: REML estimates of fixed effects for within-person covariates

Variable		Coefficients			
		VERBAL	PHYSICAL	OPS	CONTEXT
Self-efficacy	Estimate	0.180	0.210	0.270	0.100
	<i>p</i>	0.001	0.000	0.001	0.003
Prove GO	Estimate	0.010	0.040	0.070	0.020
	<i>p</i>	0.860	0.180	0.020	0.150
Avoid GO	Estimate	-0.027	-0.048	-0.045	-0.052
	<i>p</i>	0.410	0.150	0.260	0.040
Social Potency	Estimate	0.196	0.280	0.228	0.169
	<i>p</i>	0.006	0.001	0.006	0.003
Extraversion	Estimate	0.132	0.218	0.195	0.151
	<i>p</i>	0.040	0.001	0.001	0.000
Neuroticism	Estimate	-0.177	-0.081	-0.166	-0.164
	<i>p</i>	0.005	0.226	0.010	0.000
Conscientiousness	Estimate	0.308	0.161	0.207	0.325
	<i>p</i>	0.003	0.105	0.020	0.000
Agreeableness	Estimate	0.100	0.057	0.128	0.222
	<i>p</i>	0.219	0.518	0.071	0.000
Openness	Estimate	0.167	0.162	0.175	0.127
	<i>p</i>	0.012	0.012	0.005	0.001
Negative Affect	Estimate	-0.163	-0.113	-0.115	-0.080
	<i>p</i>	0.009	0.070	0.111	0.050

Table 5.13: Likelihood ratio tests for hypothesis between-person predictors

Model	$-2 \times \text{Log-likelihood}$	$\Delta\chi^2$	$p$
<hr/> Wonderlic <hr/>			
full	3369.8		
no slopes	3371.7	1.9	0.59
no int or slp	3553.1	183.3	0
<hr/> Extraversion <hr/>			
full	3373.9		
no slopes	3374.8	0.9	0.83
no int or slp	3516.6	142.7	0
<hr/> Neuroticism <hr/>			
full	3390.7		
no slopes	3392.3	1.6	0.66
no int or slp	3530.6	139.9	0
<hr/> Conscientiousness <hr/>			
full	3367.2		
no slopes	3369.3	2.1	0.55
no int or slp	3483	115.8	0
<hr/> Agreeableness <hr/>			
full	3380.6		
no slopes	3383.1	2.5	0.48
no int or slp	3527.3	146.7	0
<hr/> Openness <hr/>			
full	3377.3		
no slopes	3378.7	1.4	0.71
no int or slp	3533.1	155.8	0

Table 5.14: REML estimates of fixed effects for between-person predictors

		Coefficients				
Variable		Verb Int	Phys Int	Ops Int	Cont Int	Slope
Cognitive ability	estimate	0.015	-0.009	-0.013	-0.016	0.003
	<i>p</i>	0.111	0.280	0.159	0.002	0.117
Extraversion	estimate	0.151	0.244	0.227	0.178	-0.180
	<i>p</i>	0.018	0.001	0.003	0.000	0.268
Neuroticism	estimate	-0.162	-0.038	-0.148	-0.157	-0.014
	<i>p</i>	0.000	0.590	0.050	0.000	0.390
Conscientiousness	estimate	0.253	0.081	0.120	0.287	0.038
	<i>p</i>	0.013	0.415	0.198	0.000	0.094
Agreeableness	estimate	0.101	0.062	0.134	0.255	-0.107
	<i>p</i>	0.186	0.511	0.123	0.000	0.611
Openness	estimate	0.203	0.213	0.272	0.196	-0.038
	<i>p</i>	0.003	0.005	0.001	0.000	0.011

Table 5.15: Correlations for intercepts and slopes predicting later outcomes

Variable	Absent Hrs	State Exam	Firearms Exam	
Exam Intercept	-0.09	0.36	0.34	
Exam Slope	-0.03	-0.12	-0.15	
Self Rating Intercept	0.07	0.03	0.08	
Self Rating Slope	-0.11	0.07	0.01	
Instructor Rating Intercept	-0.06	0.12	0.27	
Instructor Rating Slope	-0.05	0.13	0.27	
Variable	Firing Score	Training Range	Award	Final Inst Rating
Exam Intercept	0.09	0.02	0.08	0.49
Exam Slope	-0.01	-0.04	-0.05	-0.37
Self Rating Intercept	0.16	-0.04	-0.08	0.17
Self Rating Slope	-0.09	0.04	0.02	0.06
Instructor Rating Intercept	0.21	0.05	0.19	0.62
Instructor Rating Slope	0.22	0.05	0.21	0.62

Table 5.16: Correlations between self- and instructor ratings of performance

Self Ratings	Instructor Ratings							
	ISELFD1	ISELFD2	ISELFA1	ISELFA2	ISELFA3	ISELFH1	ISELFH2	ISELFR1
selfdepend1	0.25	0.24	0.25	0.31	0.19	0.24	0.23	0.15
selfdepend2	0.12	0.12	0.07	0.10	0.16	0.11	0.05	0.03
selfappear1	0.00	-0.01	0.11	0.04	0.02	0.01	0.04	0.02
selfappear2	-0.06	-0.05	-0.03	-0.05	0.06	-0.01	-0.03	-0.05
selfappear3	-0.06	-0.06	-0.01	-0.05	0.17	-0.06	0.00	0.02
selfrelations1	-0.08	-0.07	-0.05	-0.05	0.02	0.04	0.03	0.05
selfrelations2	-0.10	-0.11	-0.09	-0.06	-0.03	-0.02	-0.03	-0.05
SELFR1	-0.10	-0.11	-0.09	-0.06	-0.05	-0.05	-0.10	-0.06
SELFR2	0.00	-0.02	0.05	0.04	0.02	0.03	0.07	0.03
SELFR3	-0.04	-0.03	-0.02	-0.02	0.02	-0.01	-0.01	-0.02
SELFR4	0.05	0.06	0.06	0.07	0.09	0.13	0.06	0.13
selfacademics	0.03	0.02	-0.06	0.06	0.09	0.03	0.10	0.07
selffirearms	-0.05	-0.05	-0.04	-0.02	0.05	-0.06	-0.03	-0.06
selftactics	0.09	0.09	0.09	0.14	0.11	0.17	0.10	0.11
selfscenarios	-0.06	-0.08	-0.08	-0.06	0.04	0.00	0.00	0.08
selfreports	-0.02	-0.01	-0.05	-0.06	0.03	-0.05	-0.01	0.00
	ISELFR2	ISELFR3	ISELFR4	PERF1I	PERF2I	PERF3I	PERF6I	PERF7
selfdepend1	0.25	0.24	0.15	0.16	0.01	0.10	0.03	-0.01
selfdepend2	0.03	0.14	0.09	0.18	0.00	0.06	0.02	0.09
selfappear1	-0.02	0.06	0.04	0.00	0.02	0.03	-0.11	-0.04
selfappear2	0.00	-0.02	-0.04	0.01	-0.05	-0.09	-0.10	0.02
selfappear3	-0.01	-0.02	-0.04	-0.04	-0.09	-0.10	-0.02	0.07
selfrelations1	-0.05	0.01	0.03	-0.10	-0.17	-0.03	-0.09	-0.04
selfrelations2	-0.04	-0.02	-0.02	0.04	-0.09	-0.03	-0.10	-0.06
SELFR1	-0.06	-0.06	-0.08	0.01	0.08	-0.01	-0.09	-0.09
SELFR2	0.20	-0.02	0.00	-0.14	0.01	-0.04	0.00	-0.05
SELFR3	0.00	0.03	-0.02	-0.13	-0.04	-0.12	-0.23	-0.15
SELFR4	0.15	0.12	0.15	-0.02	0.03	0.05	-0.11	0.00
selfacademics	0.00	0.00	0.02	0.30	0.10	0.13	0.10	0.12
selffirearms	0.05	-0.04	-0.02	-0.11	0.35	0.14	-0.01	0.00
selftactics	0.20	0.15	0.15	-0.02	0.14	0.08	0.10	-0.05
selfscenarios	0.02	0.02	0.01	0.02	0.13	0.10	0.11	0.09
selfreports	-0.04	-0.02	0.00	0.14	0.03	-0.01	0.00	0.13

## Chapter 6

### DISCUSSION

This study aimed to explore the multivariate and dynamic aspects of the criterion space. A definition of job performance as a behavioral repertoire was developed. This repertoire was investigated using self- and instructor-ratings of performance in a police academy setting, and examination scores. Changes in different facets of performance were mapped onto various individual differences and time-varying regulatory mechanisms.

First, confirmatory factor analyses were fit to the performance data. These analyses generally supported a bifactor model, with a general factor and several primary factors. This model marginally fit the data in the first three measurement occasions, but had problems in the final measurement occasion. Exploratory item cluster and factor analyses suggested that the performance indicators used do reflect a multifaceted construct. A general factor could be extracted in the bifactor models, but several indicators had quite low loadings on it. Furthermore, the item clusters are meaningful, showing important differentiation over time, from a strong general performance cluster through to task and contextual performance clusters and finally differentiating instructor versus self-rated task performance and self-rated contextual performance. These cluster and factor analysis results are consistent with hypothesis 1, that multiple factors would underlie the performance indicators.

The temporal patterns for the performance indicators showed generally small mean trends for various performance indicators. Exams and instructor-rated academic performance both showed small negative trends, with no random effects. All self-rated performance indicators and instructor-ratings of control tactics showed small positive trends with random effects, indicating individual differences in these mean trajectories. Finally, instructor ratings of firearms performance showed no mean trend, but a small random effect, indicating individual variability.

It was expected that corresposive effects would be found, such that the personal characteristics that led people to become police officers would display the greatest changes in response to the training context. The recruits in the sample were more extraverted, conscientious, and agreeable than typical, and less neurotic and open than typical. Furthermore, it was *a priori* expected that social potency would particularly increase and that stress reactance would particularly decrease. Social Potency, Agreeableness, and Openness showed small increases, though Social Potency showed the largest slope. Furthermore, Neuroticism and Stress Reactance displayed negative trends. There was no random effect for Stress Reactance, indicating that the decrease was essentially uniform for recruits in the sample, which is consistent with a socialization effect. Conscientiousness and Extraversion showed no significant mean time trend, but did show random slopes. This is more consistent with individual differences in change patterns than with a general contextual effect.

Hypotheses 2 and 3 examined whether different facets of performance show different patterns of change, and whether there are individual differences in change. Both were supported. Different growth curves were identified for each of the performance facets, ranging from zero for contextual performance through the lower .20s for operations performance. The likelihood ratio test that these facets do not show different growth patterns would have been impossible within a univariate framework. Furthermore, individual differences in the intercepts were found for these facets and there were small but significant individual differences in the slopes, as well. Some of this variability could be explained by level-2 Openness and Conscientiousness, though surprisingly, not cognitive ability. Cognitive ability was only a significant and negative predictor of the intercept for contextual performance.

Residual within-measurement occasion variance, controlling for the growth pattern, could be explained using time-varying covariates. These latter findings provide some support for hypotheses 4 and 5, though not terribly strong evidence, as the expected pattern was not supported. For instance, it was not expected that Agreeableness and Openness would show changes comparable to traits such as Extraversion and Neuroticism which were expected to change during training. Additionally, the variability over time in these traits was also associated with performance outcomes, which was unexpected. These findings do call into question the theoretical logic of the study, or at least suggest

that the theoretical model requires some additional explanations.

Moving to more traditional univariate approaches for exam scores and instructor ratings of performance showed patterns of results that were more consistent with previous research. Cognitive ability, Extraversion, and Agreeableness emerged as individual difference predictors of exam score intercepts, though no evidence was found for a random slope for the growth trend on exams. The only time-varying covariate that was associated with exam scores was performance-prove goal orientation, which showed a substantial negative relationship.

Regarding the instructor ratings, all conclusions should be regarded somewhat cautiously. There was very little variance observed in these variables. There was no general time trend in this outcome, but there was substantial within-person variability, as illustrated in the trellis plots. Cognitive ability again emerged as a level-2 predictor of the intercept, though its coefficient was somewhat small. Agreeableness emerged as a negative predictor of ratings' intercept, perhaps because overly-agreeable individuals are perceived by instructors as too weak or kind for police work. Performance-prove orientation again emerged as a level-1 predictor, along with self-efficacy and negative affect, all of which showed random slopes.

Finally, the ability of the trajectories of individual random intercept and slope coefficients for “knowledge”, as indexed by exam scores, and “skill”, as indexed by self- and instructor-ratings of simulation performance, as predictors of later training outcomes was examined. Exam score intercepts were useful in predicting passing the State certification exam and Firearms exam scores and final overall instructor ratings. Exam slopes negatively predicted final instructor ratings, but this is possibly a ceiling effect or a byproduct of the extremely limited variability around the general slope on time for exam scores. The intercepts for instructor ratings were somewhat predictive of passing the State certification exam and receiving an award, and moderately predictive of Firearms exam scores and firing range ratings. Instructor rating intercepts were, naturally, highly predictive of final instructor ratings. The trajectories of instructor ratings were so highly related to intercepts that they provided no unique information.

The major hypotheses derived from the theoretical model were generally supported. However, the supplementary expectations that would help to confirm whether or not the socialization mechanisms postulated were re-

sponsible for the observed changes were only partially supported. Therefore, it is best to conclude that the job performance repertoire likely develops in the same way that underlies the general adult development process, including such things as personality change and maturity.

## 6.1 Limitations

The empirical study reported is correlational in nature. Therefore, the most damaging limitation of the work is that no strong causal inferences can be made. It was not possible to obtain a waitlist-style control group, so it is not possible to strongly infer that the changes observed during this study are attributable to the training program. However, this limitation is likely to occur in any realistic prospective validation design, and so reflects a real practical constraint on inference.

Fortunately, the study does employ a prospective longitudinal design, with all variables (except cognitive ability) measured at every time point. As a result, it is possible to test cross-lagged effects and to therefore test directional hypotheses.<sup>7</sup> Additionally, including mediator variables helps to clarify the potential causal sequence. Unfortunately the pattern of mediation was not supported, so this design feature did not provide strong inferences about the causal pathways. Some support for the measured individual-differences serving as cross-level moderators was found, but this evidence was not replicable over performance facets or time-varying covariates.

This leaves us with the one potentially damning limitation of the study reported here. The time sequence for measurement was chosen largely for convenience. There is little in the way of theory to suggest when to make repeated criterion measurements. In fact, a strength of this study is that it may speak to what are and are not appropriate time lags between measurements.

A further limitation is that the sample is composed entirely of police personnel. The predictors used are generally accepted as being valid over jobs and performance criteria (e.g., Schmidt & Hunter, 1998; Barrick & Mount, 1991). Additionally, the criterion measures are based on the Campbell model of performance (Campbell, 1990) and have been shown to be broadly appli-

---

<sup>7</sup>An initial Granger causality test was significant for Self-efficacy  $\rightarrow$  Exams ( $p < .0001$ ) and not for Exams  $\rightarrow$  Self-efficacy ( $p = .14$  Granger, 1969).



cable over a wide variety of jobs (e.g., Campbell, McHenry, & Wise, 1990). These dimensions have such broad applicability because they are relatively abstract; they allow for considerable specialization in the particular performance indicators used to measure them. Training criteria measured include knowledge retention and behavior skill demonstration, which are very general types of training outcomes (e.g., Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997).

## 6.2 Future Directions

This study addresses the dynamic influence of individual differences on a wide array of job performance behaviors. However, they leave open a variety of questions regarding the pathways of influence. Data collected here begin to address the mediation of intelligence and personality traits by more context-specific variables, which builds on work by Schmidt and Hunter (1992; Schmidt, Hunter, & Outerbridge, 1986) and integrates aspects of the Neo-Socioanalytic Model of personality.

However, there is considerable room open to investigate the processes by which these variables wield their influence. Several theoretical models exist that can be investigated, such as Klein's (1989) control theoretic model of motivation or Beal, Weiss, Barros, and MacDermid's (2005) Episodic Process Model, which discusses the effects of attentional and affective regulation in discrete performance episodes. Both involve the self-regulation of behavior at work over time, and provide theoretical insights into how individual differences and self-regulation might be fruitfully integrated to study work behavior, including performance (e.g., Cervone, Shadel, Smith, & Fiori, 2006). Furthermore, the Neo-Socioanalytic Model of personality can again provide a useful framework for studies integrating these different approaches (Bogg et al., 2008).

Future work could combine studies of work behavior across several time scales. For instance, ecological momentary sampling studies of contextual performance and counter-productive work behaviors could be compared to longer term studies of the same performance facets, based on annual or semi-annual performance evaluation studies. Such studies could help to clarify the time sequencing of behavior. It is even possible that different mechanisms

are responsible for short-term variability in performance behaviors than for long-term development.

## 6.3 Conclusions

This study allows a new perspective on the validation of a selection battery. These data also provide insight into the operation of a variety of individual differences at work in training. This study examined the prediction of performance across several facets of the criterion domain over time. This allowed a powerful prediction model of the individual differences-performance relationship to be tested. The information collected allowed for testing the mediating pathways of a variety of affective and regulatory mechanisms. The design presented here opens up an entirely new framework for the validation of selection tools. This design provides information about all sources of lawful variance in the job performance domain (Ghiselli, 1956; Inn et al., 1972) and allows researchers and practitioners to understand how the predictors used operate over time.

## Appendix A: Measures

### Work Goal Orientation (Vandewalle, 1997)

Please evaluate yourself on the following statements. The scale reflects your agreement with how well the statement describes you. 1=Not at all 2=Slightly agree 3= Neither agree nor disagree 4 = Moderately agree 5 = Strongly Agree

1. I often read materials related to my work to improve my ability.
2. I am willing to select a challenging work assignment that I can learn a lot from.
3. I often look for opportunities to develop new skills and knowledge.
4. I enjoy challenging and difficulty tasks at work where I'll develop new skills.
5. For me, development of my work ability is important enough to take risks.
6. I prefer to work in situations that require a high level of ability and talent.
7. I'm concerned with showing that I can perform better than my coworkers.
8. I try to figure out what it takes to prove my ability to others at work.
9. I enjoy it when others at work are aware of how well I'm doing.
10. I prefer to work on projects where I can prove my ability to others.
11. I would avoid taking on a new task if there was a chance I would appear rather incompetent to others.
12. Avoiding a show of low ability is more important to me than learning a new skill.
13. I'm concerned with taking on a new task at work if my performance might show I had low ability.

14. I prefer to avoid situations at work where I might perform poorly.

### Job-specific Self Efficacy (Jones, 1986)

Please evaluate yourself on the following statements. The scale reflects your agreement with how well the statement describes you. 1 = Not at all 2 = Slightly agree 3 = Neither agree nor disagree 4 = Moderately agree 5 = Strongly Agree

1. My new job is well within the scope of my abilities.
2. I do not anticipate any problems in adjusting to work in this organization.
3. I feel I am overqualified for the job I will be doing.
4. I have all the technical knowledge I need to deal with my new job, all I need now is practical experience.
5. I feel confident that my skills and abilities equal or exceed those of my future colleagues.
6. My past experiences and accomplishments increase my confidence that I will be able to perform successfully in this organization.
7. I could have handled a more challenging job than the one I will be doing.
8. Professionally speaking, my new job exactly satisfies my expectations of myself. (R)

## RECRUIT PERFORMANCE SELF-EVALUATION FORM

Answer the following questions. The aim of this evaluation is to take a critical look at your training, to identify those areas you are doing well, and/or those that you need improvement. (Note: All items except for Teamwork and Citizenship also rated by Instructors)

1 = Poor, 2 = Needs Improvement, 3 = Average, 4 = Above Average, 5 = Excellent

### DEPENDABILITY:

1. I am punctual in arriving at activities, returning from breaks, and complying with curfew.
2. I turn in assignments on time.

### APPEARANCE/DEMEANOR/MAINTENANCE:

1. I wear my uniform in accordance with my department policy.
2. I am professional in my demeanor.
3. My quarters are well maintained.

### HUMAN RELATIONS SKILLS:

1. I interact appropriately with my classmates.
2. I interact professionally with others.

### TEAMWORK AND CITIZENSHIP

1. I assist classmates with training exercises as needed.
2. I assist instructors when asked.
3. I take initiative and volunteer for scenario exercises.

### RESPONSE TO TRAINING:

1. I maintain a positive attitude when receiving instructions or suggestions.
2. I take the initiative whenever possible and assume a leadership role.
3. I use the lessons I've learned from my peers and/or instructors.
4. I am involved in the training.

Area of Training	Poor Improvement	Needs	Competent, but needs continued work	Competent	Exceeds Expectations
Academics	1	2	3	4	5
Firearms	1	2	3	4	5
Control Tactics	1	2	3	4	5
Patrol Operations	1	2	3	4	5
Vehicle Stops	1	2	3	4	5
Integrated Scenarios	1	2	3	4	5
Report-Writing	1	2	3	4	5

## REFERENCES

- Ackerman, P.L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102, 3 - 27.
- Alliger, G.M, Tannenbaum, S.I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta- analysis of the relations among training criteria. *Personnel Psychology*, 50, 341-358.
- Alvares, K.M., & Hulin, C.L. (1972). Two explanations of temporal changes in ability-skill relationships: A literature review and theoretical analysis. *Human Factors*, 14, 295-308.
- Alvares, K.M., & Hulin, C.L. (1973). An experimental evaluation of a temporal decay in the prediction of performance. *Organizational Behavior and Human Performance*, 9, 169 - 185.
- Angoff, W.H. (1988). Validity: An evolving concept. In H. Wainer and H. Braun (Eds.). *Test validity* (pp. 19-32). Hillsdale, NJ: Lawrence Erlbaum.
- Astin, A. (1964). Criterion-centered research. *Educational and Psychological Measurement*, 24, 807-822.
- Austin, J.T., Humphreys, L.G., & Hulin, C.L. (1989). Another view of dynamic criteria: A critical reanalysis of Barrett, Caldwell, & Alexander. *Personnel Psychology*, 42, 583 - 596.
- Austin, J.T., & Villanova, P. (1992). The criterion problem: 1917 - 1992. *Journal of Applied Psychology*, 77, 836 - 874.
- Baltes, P.B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, 3, 611 - 627.
- Baltes, P.B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. *American Psychologist*, 366 - 380.



- Barrett, G.V., Miguel, R.F., Hurd, J.M., Lueke, S.B., & Tan, J.A. (2003). Practical issues in the use of personality tests in police selection. *Public Personnel Management*, 32, 497 - 517.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M.R., Mount, M.K., & Strauss, J.P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology*, 78, 715 - 722.
- Barrick, M.R., Mount, M.K., & Stewart, G.L. (2003). Situational and motivational influences on trait-behavior relationships. In M.R. Barrick & A. M. Ryan (Eds.) *Personality and work: Reconsidering the role of personality in organizations*. (pp. 60 - 82). San Francisco, CA: John Wiley & Sons.
- Bashshur, M.R. (2006). *Modeling individual and environmental influences on performance determinants: The search for linear, quadratic, and multiplicative effects*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Beal, D.J., Weiss, H.M., Barros, E., & MacDermid, S.M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology*, 90, 1054-1068.
- Bliese, P. D. (2002). *Multilevel Modeling in R: A brief introduction to R, the multilevel package, and the NLME package*. Walter Reed Army Institute of Research.
- Bliese, P. D. & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods*, 7, 400-417.
- Bliese, P.D., & Ployhart, R.E. (2002). Growth modeling using random coefficient models: Model building, testing and illustrations. *Organizational Research Methods*, 5, 362-388.
- Bogg, T., Voss, M.W., Wood, D., & Roberts, B.W. (2008). A hierarchical investigation of personality and behavior: Examining Neo-Socioanalytic models of health-related outcomes. *Journal of Research in Personality*, 42, 183-207.
- Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M. Dunnette & L. Hough (Eds.) *Handbook of Industrial/Organizational Psychology*, (Vol. 2, 271-326). Palo Alto, CA: Consulting Psychologists Press.

- Borman, W.C., Hanson, M.A., & Hedge, J.W. (1997). Personnel selection. *Annual Review of Psychology*, 48, 299-337.
- 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., Pulakos, E.D., & Oppler, S.H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology*, 76, 863 - 872.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. London: Sage.
- 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* (Vol. 1, 2nd ed.; pp. 687-732). Palo Alto, CA: Consulting Psychologists Press, Inc.
- 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.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations*. San Francisco: Jossey-Bass.
- Campbell, J. P., McHenry, J.J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology*, 43, 313-333.
- Carver, C.S., & Scheier, M.F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97, 19 - 35.
- Cervone, D., Shadel, W.G., Smith, R.E., & Fiori, M. (2006). Self-regulation: Reminders and suggestions from personality science. *Applied Psychology: An International Review*, 55, 333-385.
- Chan, D. (1998). The conceptualization and analysis of change over time: A integrative approach incorporating longitudinal mean and covariance structures analysis (LMACs) and multiple indicator latent growth model (MLGM). *Organizational Research Methods*, 14, 421-483.
- Chan, K.-Y., & Drasgow, F. (2001). Toward a theory of individual differences and leadership: Understanding the motivation to lead. *Journal of Applied Psychology*, 86, 481-498.

- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*, 678-707.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Dalal, R. S. & Hulin, C. L. (2008). Motivation in organizations: Criteria and dynamics. In R. Kanfer, G. Chen, & R. Pritchard (Eds.), *Work Motivation: Past, Present, and Future* (pp. 63-100). New York: Erlbaum.
- Day, D.V., Harrison, M.M., & Halpin, S.M. (2009). *An integrated approach to leader development: Connecting adult development, identity, and expertise*. New York: Routledge.
- Day, D.V., & Sin, H.P. (2009). Leader development, identity, and goal orientation: A study of personal change trajectories. *Proceedings of the 2009 Annual Meeting of the Academy of Management (CD)*.
- Deadrick, D.L., Bennett, N., & Russell, C.J. (1997). Using hierarchical linear modeling to examine dynamic performance criteria over time. *Journal of Management, 23*, 745-757.
- Donnellan, M.B., Conger, R.D., & Burzette, B.G. (2005). Criterion-related validity, self-other agreement, and longitudinal analyses for the Iowa Personality Questionnaire: A short alternative to the MPQ. *Journal of Research in Personality, 39*, 458- 485.
- Dunnette, M.D. (1963). A note on the criterion. *Journal of Applied Psychology, 47*, 251-254.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: A guide to theory and research*. Reading, MA: Addison-Wesley.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*, 1011-1027.
- Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modeling, 13*, 465 - 486.
- Frederickson, B.L. (2001). The role of positive emotions in positive psychology: The broaden and build theory of positive emotions. *American Psychologist, 56*, 218-226.

- Gellatly, I.R. (1996). Conscientiousness and task performance: Test of a cognitive-process model. *Journal of Applied Psychology*, 81, 474-482.
- Ghiselli, E.E. (1956). Dimensional problems of criteria. *Journal of Applied Psychology*, 40, 1-4.
- Ghiselli, E. E., & Haire, M. (1960). The validation of selection tests in the light of the dynamic nature of the criteria. *Personnel Psychology*, 13, 225-231.
- Hanisch, K.A., Hulin, C.L., & Roznowski, M. (1998). The importance of individuals' repertoires of behaviors: The scientific appropriateness of studying multiple behaviors and general attitudes. *Journal of Organizational Behavior*, 19, 463-480.
- Heckman, J.J. (2008). Econometric causality. *International Statistical Review*, 76, 1 - 24.
- Hofmann, D.A., Jacobs, R., & Baratta, J.E. (1993). Dynamic criteria and the measurement of change. *Journal of Applied Psychology*, 78, 194-204.
- Hoffman, L., & Rovine, M.J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39, 101 - 117.
- Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88, 100-112.
- Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist*, 46, 469 - 477.
- Hogan, R., & Roberts, B.W. (2004). A socioanalytic model of maturity. *Journal of Career Assessment*, 12, 207-217.
- Hollenbeck, J.R., Williams, C.R., & Klein, H.L. (1989). An empirical examination of the antecedents to difficult goals. *Journal of Applied Psychology*, 74, 18-23.
- Hulin, C.L. (1982). Some reflections on general performance dimensions and halo rating error. *Journal of Applied Psychology*, 67, 165-170.
- Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability and predictive relationships. *Psychological Bulletin*, 107, 328-340.
- Humphreys, L.G. (1960). Investigations of the simplex. *Psychometrika*, 25, 313-323.

- Hunter, J.E. (1983). A causal analysis of cognitive ability, job knowledge, job performance and supervisory ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257 - 266). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Inn, A., Hulin, C.L., & Tucker, L.R. (1972). Three sources of criterion variance: Static dimensionality, dynamic dimensionality, and individual dimensionality. *Organizational Behavior and Human Performance*, 8, 58-83.
- John, O.P., & Srivastava, S. (1999). The big-five trait taxonomy: History, measurement, and theoretical perspectives. In L. Pervin & O.P. John (Eds.). *Handbook of personality: Theory and research* (2nd ed., pp. 102 - 138). New York: Guilford.
- Johnson, J.W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology*, 86, 984-996.
- Johnson, J.W. (2003). Toward a better understanding of the relationship between personality and job performance. In M. Barrick & A.M. Ryan (Eds.) *Personality and Work* (pp. 83-120). San Francisco, CA: John Wiley & Sons.
- Jones, G.R. (1986). Socialization tactics, self-efficacy, and newcomers' adjustments to organizations. *Academy of Management Journal*, 29, 262 - 279.
- Judge, T.A., Erez, A., Bono, J.E., & Thoreson, C.J. (2002). Do the traits self-esteem, neuroticism, locus of control, and generalized self-efficacy indicate a common core construct? *Journal of Personality and Social Psychology*, 83, 693-710.
- Judge, T.A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta- analytic review. *Journal of Applied Psychology*, 87, 797-807.
- Kane, J.S. (1986). Performance distribution assessment. In R.A. Berk (Ed.), *Performance assessment: Methods and applications* (pp. 237-273). Baltimore: Johns Hopkins University Press.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An Integrative/aptitude-treatment approach to skill acquisition. *Journal of Applied Psychology*, 74, 657-690.

- Keil, C. T., & Cortina, J. M. (2001). Degradation of validity over time: A test and extension of /bin/bash: :wq: command not found
- Kerr, S. (1995). An Academy classic. On the folly of rewarding A, while hoping for B. *Academy of Management Executive*, 9, 7 - 14.
- Klein, H.J. (1989). An integrated control theory model of work motivation. *Academy of Management Review*, 14, 150-172.
- Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86, 148-161.
- Lee, S., & Klein, H.J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology*, 87, 1175-1182.
- Lodi-Smith, J., & Roberts, B.W. (2007). Social Investment and Personality: A Meta-Analysis of the Relationship of Personality Traits to Investment in Work, Family, Religion, and Volunteerism. *Personality and Social Psychology Review*, 11, 68 - 86.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83-104.
- Martocchio, J., & Judge, T.A. (1997). Relationship between conscientiousness and learning in employee training: Mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology*, 82, 764-773.
- McCrae, R.R., & Costa, P.T. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J.S. Wiggins (Ed.), *The five-factor model of personality* (pp. 51 - 87). New York: Guilford.
- Messick, S. (1989). Validity. In R. Linn (Ed.) *Educational measurement* (3rd ed.). The American Council on Education/Macmillan series on higher education. (pp. 13-103). New York, NY: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741 - 749.
- Motowidlo, S.J., Borman, W.C., & Schmit, M.J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10, 71-84.

- 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.
- Murphy, K.R. (1989). Is the relationship between cognitive ability and performance be stable over time? *Human Performance*, 2, 183-200.
- Murphy, K.R. (2009). Validity, Validation, and Values. *The Academy of Management Annals*, 3, 421 - 461.
- Murphy, K.R. (In press). How a broader definition of the criterion domain changes our thinking about adverse impact. In J. Outtz (Ed). *Adverse Impact*. San Francisco, CA: Jossey-Bass.
- 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.
- Ones, D. S. (1993). *The construct validity of integrity tests*. Unpublished doctoral dissertation, University of Iowa, Iowa City.
- Ones, D.S., Mount, M.K, Barrick, M.R., & Hunter, J.E. (1994). Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. *Personnel Psychology*, 47, 147-156.
- Patrick, C.J., Curtain, J.J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*, 14, 150-163.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. (2nd Ed.) Cambridge, UK: Cambridge University Press.
- Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.
- Ployhart, R.E., & Hakel, M.D. (1998). The substantive nature of performance variability: Predicting interindividual differences in intraindividual performance. *Personnel Psychology*, 51, 859 - 901.
- Pulakos, E.D., Arad, S., Donovan, M.A., & Plamondon, K.E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, 85, 612-624.
- R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria
- Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1, 86-89.

- Revelle, W. (1979). Hierarchical cluster-analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14, 57-74.
- Revelle, W. (2010). *psych: Procedures for personality and psychology research*. R package version 1.0-86.
- Roberts, B.W. (2006). Personality development and organizational behavior. In B.M. Staw (Ed). *Research in organizational behavior* (pp. 1-41). Elsevier Science/JAI Press.
- Roberts, B.W., & Caspi, A. (2003). The cumulative continuity model of personality development: Striking a balance between continuity and change in personality traits across the life course. U.M Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 183-214). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Roberts, B.W., & Robins, R.W. (2000). Broad dispositions, broad aspirations: The intersection of personality traits and major life goals. *Personality and Social Psychology Bulletin*, 26, 1284 - 1296.
- Roberts, B.W., Walton, K.E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132, 1-25.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications from 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work samples performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, 71, 432-439.
- Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40, 437-453.
- Schneider, R.J., Hough, L.M., & Dunnette, M.D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior*, 17, 639 - 655.
- Singer, J.D. (1998). Using SAS PROC MIXED to fit multilevel, hierarchical, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23, 323-355.
- Smith, P.C. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 745 - 775). Chicago: Rand McNally.



- Sobel, M.E. (1982). Asymptotic confidence intervals for the indirect effect in structural equations models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290-312). Washington, DC: American Sociological Association.
- Spain, S.M., Miner, A.G., Kroonenberg, P.M., & Drasgow, F. (In press). Job performance as multidimensional dynamic criteria: Experience sampling and multiway component analysis. In press at *Multivariate Behavioral Research*.
- Srivastava, S., John, O.P., Gosling, S.D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*, 1041-1053.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703-742.
- Vandewalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement*, *57*, 995 - 1015.
- Wernimont, P.F., & Campbell, J.P. (1968). Signs, Samples, and Criteria. *Journal of Applied Psychology*, *52*, 372 - 376.
- Weitz, J. (1960). Criteria for criteria. *American Psychologist*, *16*, 228 - 231.
- Winter, D.G., John, O.P., Stewart, A.J., Klohnen, E.C., & Duncan, L.E. (1998). Traits and motives: Towards and integration of two traditions in personality research. *Psychological Review*, *105*, 230-250.
- Wood, D., & Roberts, B.W. (2006). Cross-sectional and longitudinal tests of the Personality and Role Identity Structural Model (PRISM). *Journal of Personality*, *74*, 779-809.
- Wonderlic, Inc. (1999). *Wonderlic Personnel Test*. Wonderlic & Associates: Libertyville, IL.
- Watson, D.A., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063 - 1070.
- Zyphur, M.J., Bradley, J.C., Landis, R.S., & Thoresen, C.J. (2008). The effects of cognitive ability and conscientiousness on performance over time: A censored latent growth model. *Human Performance*, *21*, 1 - 27.

# VITA

## Seth M Spain

University of Nebraska

Department of Management

College of Business, CBA 214

Lincoln, NE 68588-0405

Phone: (217) 778-0011

Email: [smSpain@gmail.com](mailto:smSpain@gmail.com)

### Education

PhD. Industrial and Organizational Psychology, University of Illinois at Urbana-Champaign, 2010.

M.A. Industrial and Organizational Psychology, University of Illinois at Urbana-Champaign, 2007.

B.S. Psychology, University of Illinois at Urbana-Champaign, 2003.

### Employment

Postdoctoral Research Fellow, University of Nebraska at Lincoln, 2010–present.

Internal Consultant, State Universities Civil Service System, 2005–2010.

### Research Interests

Organizational Behavior and Human Resource Management

Dynamic Criteria

Personality and Affect

Research Methods

## Publications

### Articles in press

Spain, S.M., Miner, A.G., Kroonenberg, P.M., & Drasgow, F. (In press). Job performance as multivariate dynamic criteria: Experience sampling and multiway component analysis. *Multivariate Behavioral Research*.

### Journal Articles

Brummel, B.J., Rupp, D.E., & Spain, S.M. (2009). Constructing parallel simulation exercises for assessment centers and other forms of behavioral assessment. *Personnel Psychology*, 62, 137–170.

Gibbons, A.M., Rupp, D.E., Baldwin, A.M., Snyder, L.A., Spain, S.M., Woo, S.E., Brummel, B.J., Sims, C.S., & Kim, M.-J. (2006). Initial validation of developmental assessment center. *The Psychologist-Manager Journal*, 9, 75–98.

### Manuscripts under review

Spain, S.M., Jackson, J.J., & Edmonds, G.W. Extending the Actor-Partner Interdependence Model to Include Binary Outcomes: A multilevel logistic approach. Submitted to *Personal Relationships*.

Milanak, M.E., Berenbaum, H., Spencer-Smith, J., & Spain, S.M. Exploring models of facial affect recognition: The roles of intensity and reaction time. Revision at *Cognition and Emotion*.

## Presentations and Posters

Gibbons, A.M., Spain, S.M., & Vanhove, A. (2010, April). Describing Inconsistent Assessment Center Ratings: Simplex Models of Exercise Similarity. In D. Jackson & B. Hoffman (Chairs). Exercise-driven Variance in Assessment Center Ratings: Alternative Approaches, New Insights. Symposium presented at the 25th annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA, April 8-10.

Newman, D.A., Spain, S.M., Joseph, D.L., Fisher, C.D., Glomb, T.M., & Miner, A.G. (2010, April). Intrinsic dynamic regulation of work satisfaction and mood. In P. Hanges & A. Fulmer (Chairs). New Developments in Modeling Longitudinal and Dynamic Data. Symposium presented at the 25th annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA, April 8-10.

Spain, S.M., Tay, & L., Diener, E. (2009, January). Exploring the dimensions of affective experiences with three-mode component analysis. Poster presented at the 2009 annual meeting of the Society for Social and Personality Psychology, Las Vegas, NV, Jan 28-30, 2010.

Spain, S.M. (2009, August). It's about time (and change). (Chair) Paper session presented at the 2009 annual meeting of the Academy of Management, Chicago, IL, August 7-11.

Spain, S.M. (2009, April). Multivariate dynamic criteria: A practical approach. In D.A. Newman (Chair), Time and Job Performance. Symposium presented at the 24th annual meeting of the Society of Industrial and Organizational Psychology, New Orleans, LA, April 2-4.

Newman, D.A., & Spain, S.M. (2009, April). A latent growth model of validity degradation. In D.A. Newman (Chair). Time and Job Performance. Symposium presented at the 24th annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA, April 2-4.

von Thaden, T.L., Spain, S.M., & Woo, S.E. (2009, April). Self-reported fatigue and organizational risk in multiple airlines. In T.L. von Thaden (Chair), Organizational Safety and Effectiveness. Symposium Presented

at the 2009 International Symposium on Aviation Psychology, Dayton, OH, April 27-30.

von Thaden, T.L., Woo, S.E., & Spain, S.M. (2009, April). Validating a four-factor model of safety culture in commercial flight operations. In T.L. von Thaden (Chair), *Organizational Safety and Effectiveness*. Symposium Presented at the 2009 International Symposium on Aviation Psychology, Dayton, OH, April 27-30.

von Thaden, T.L., Woo, S.E., & Spain, S.M. (2009, April). Investigating national differences in commercial aviation safety culture: A comparison of flight operations. In T.L. von Thaden (Chair), *Organizational Safety and Effectiveness*. Symposium Presented at the 2009 International Symposium on Aviation Psychology, Dayton, OH, April 27-30.

Spain, S.M., & Miner, A.G. (2008, August). Three-mode principal components analysis of daily affect at work. Interactive paper presented at the 2008 annual meeting of the Academy of Management, Anaheim, CA, August 3-8.

Spain, S.M., Jackson, J., Edmonds, G.W., & Roberts, B.W. (2008, May). Hierarchical polynomial regression for modeling interaction effects with dyad data. Poster presented at the 2008 annual meeting of the Association for Psychological Science, Chicago, IL, May 22-25.

Spain, S.M., Jackson, J., Edmonds, G., & Roberts, B.W. (2008, February). Modeling personality complementarity in dyads with polynomial regression and response surfaces. Poster presented at the 2008 annual meeting of the Association for Research in Personality, Albuquerque, NM, February 6-7.

Rupp, D.E., & Spain, S.M. (2007, August). Corporate social responsibility and organizational justice: Multi-foci,-level,-motive perspective. (Chair) Symposium at the 2007 annual meeting of the Academy of Management, Philadelphia, PA, August 3-8.

Spain, S.M., & Miner, A.G. (2007, August). Experience sampling and multiway analysis: The dynamic structure of job performance. Paper at

the 2007 annual meeting of the Academy of Management, Philadelphia, PA, August 3-8.

Spain, S.M., & Miner, A.G. (2007, January). Affect at work: Experience sampling and multiway analysis. Poster presented at the 2007 Emotions preconference of the 8th annual meeting of the Society for Personality and Social Psychology, Memphis, TN, January 24-27.

Brummel, B.J., & Spain, S.M. (2005, April). Constructing parallel simulation exercises for developmental assessment centers. In L. A. Snyder, & D. E. Rupp (Chairs), *Developmental Assessment Centers: Special Considerations for Researchers and Practitioners*. Symposium presented at the 20th Annual Meeting of the Society for Industrial Organizational Psychology, Los Angeles, CA, April 15-17.

### Manuscripts in preparation

Spain, S.M., & Newman, D.A. A latent growth model of validity degradation. In preparation for *Journal of Applied Psychology* (writing).

Spain, S.M., Tay, S., Diener, E., & Miner, A.G. Exploring the dimensions of affective experiences with three-mode component analysis (writing).

von Thaden, T.L., Spain, S.M., & Woo, S.E. Self-reported fatigue in multiple airlines. In preparation for *Human Performance* (analyzing data).

### Technical reports and trade publications

Spain, S.M., & Burrus, K.D. (2007). *Quantitative task comparison: A method for reviewing job descriptions and class specifications*. Technical report prepared for the State Universities Civil Service System, Urbana, IL.

Spain, S.M., Klafehn, J.L., & McCanse, A.S. (2007). A primer on classification and examination procedures. *The System News: Illinois' Public University Support Resource and Advocate*, 5 (2), 2-6 [Special issue].

## Current Projects

Police Examinations Longitudinal Validation Project, State Universities  
Civil Service System, Urbana, IL.

Evaluation of Police Training Delivery, Police Training Institute at the  
University of Illinois, Urbana-Champaign.

## Teaching experience

Teaching Assistant/Lecturer, “Introduction to Social Psychology”  
Perfect Teaching Rating (5.0/5.0), Fall 2009

Guest Lecturer, “Organizational Psychology” (Daniel Newman)

Instructor, “Introduction to Industrial/Organizational Psychology”  
Rated as Excellent, Spring 2008

Teaching Assistant, to Deborah Rupp in “Organizational Psychology”

## Teaching Interests

Personnel Selection & Performance/HR

Organizational Psychology/OB,

Research Methods (particularly, Structural Equations/Latent Growth Models,  
Multilevel models, Generalized Linear Models)

## Consulting

Police Training Institute  
External Consultant, 2008-Present

State Universities Civil Service System  
Intern, 2005-2007

Internal Consultant, 2007-Present

## Technical Skills

### Computer Skills

Windows OSes, UNIX/LINUX/CYGWIN, Mac OS X

SAS, MATLAB, SPSS, LISREL, BILOG, R (some), Stata (Novice)

HTML, Perl, Python, L<sup>A</sup>T<sub>E</sub>X, Fortran 95 (some)

MacVim/gVim is my preferred text editor

### Statistical Expertise

Multilevel modeling

Structural equation modeling

Latent growth models

Multiway models

Generalized linear mixed models

## Honors and Awards

List of Teachers Rated as Outstanding by Their Students, University of Illinois at Urbana-Champaign

List of Teachers Rated as Excellent by Their Students, University of Illinois at Urbana-Champaign

Dean's List, University of Illinois, Urbana-Champaign



## Professional

### Memberships

Academy of Management-Student Affiliate (OB, RM divisions)

Society for Industrial and Organizational Psychology-Student Affiliate

Society for Personality and Social Psychology-Student Affiliate

### Service

Membership Committee, Academy of Management

Conference Reviewer, Academy of Management (HR, OB, RM)

### Ad Hoc Reviewer

Multivariate Behavioral Research

## References

### Academic

Fritz Drasgow, Ph.D.  
Department of Psychology,  
603 E. Daniel St.,  
Champaign-Urbana, IL 61820  
Email: fdrasgow@illinois.edu  
Telephone: 217-333-2739

Charles L. Hulin, Ph.D. (Emeritus)  
Department of Psychology,  
603 E. Daniel St.,  
Champaign-Urbana, IL 61820  
Email: chulin@illinois.edu

Daniel A. Newman, Ph.D.  
Department of Psychology,  
603 E. Daniel St.,  
Champaign-Urbana, IL 61820  
Email: d5n@illinois.edu  
Telephone: 217-390-3048

### Professional

Lewis T. (Tom) Morelock  
Executive Director,  
State Universities Civil Service System  
1717 Philo Rd, Suite 24  
Urbana, IL 61802  
Email: TomM@sucss.state.il.us  
Telephone: 217-278-3150