Spring 6-14-2019

Effects of Personality Retesting on Validity Coefficients as Moderated by Situational Strength

Daniel R. Abben
DePaul University, dabben@depaul.edu

Effects of Personality Retesting on Validity Coefficients

as Moderated by Situational Strength

A Dissertation

Presented in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

By

Daniel Richard Abben

June, 2019

Department of Psychology

College of Science and Health

DePaul University

Chicago, Illinois

Dissertation Committee

Alice F. Stuhlmacher, Ph.D., Chairperson

Goran Kuljanin, Ph.D.

Kimberly Quinn, Ph.D.

Suzanne Fogel, Ph.D.

Bruno Teboul, Ph.D.

Acknowledgements

I would like to thank my committee members, especially Alice Stuhlmacher, Ph.D, and the DePaul I/O faculty, especially Jane Halpert, Ph.D. I would like to thank Filip Lievens, Ph.D., for allowing me to use the in-basket measure that was part of this study. Finally, I would like to acknowledge my family and my circle of friends, particularly Steve Cisneros, Charlie Wright, and the Arthur Avenue cluster.

Biography

The author was born in Waterloo, Iowa, on June 20, 1982. He graduated from Columbus High School, in Waterloo. He received his Bachelor of Arts degree from Clarke College, now Clarke University, in 2004. He received his Master of Science and Master of Arts degrees from DePaul University in 2011 and 2014, respectively.

Table of Contents

List of Tables

Abstract

The use of valid selection tests enables organizations to better select employees who have the knowledge, skills, abilities, and other characteristics that are necessary for success. While cognitive ability tests are one of the best predictors of performance, they have well-known limitations. Specifically, they can result in adverse impact, and there is clear evidence of retest effects. The use of personality tests, when included in a selection battery, can ameliorate adverse impact and can provide incremental validity. Personality tests, however, also have limitations. Namely, they can be faked, the construct can be measured in various ways (i.e., there are myriad constructs, many of which can be measured at a trait level and a facet level), and there are numerous moderators of the personality–performance relationship. This lab study investigated the facet-level conscientiousness–performance relationship, explored evidence regarding practice effects in personality tests, examined whether situational strength moderated the personality–performance relationship, and looked at the form of the relationship between personality and performance. A unique contribution of this study is that performance was operationalized by using an in-basket that assessed four different dimensions of performance. In general, the hypotheses were not supported, underscoring the need for future research.

Introduction

Within organizations, employee selection is an important endeavor. When the right people are hired, companies are better able to carry out their objectives effectively and efficiently. Ineffective hiring can result in a lack of person and organization fit, increased turnover, and poor performance. As such, organizations strive to use selection systems that are high in utility, reliability, and validity.

A significant body of literature suggests that measures of cognitive ability effectively predict a number of organizationally relevant outcomes, including job performance, training performance, and organizational attainment (Blume, Ford, Baldwin, & Huang, 2010; Hunter & Hunter, 1984; Lievens, Harris, Van Keer, & Bisqueret, 2003; Ree, Carretta, & Teachout, 1995; Schmidt, 2002; Schmidt & Hunter, 1998, 2004). Despite this fact, cognitive ability tests are not perfect. In addition to the evidence regarding the validity of these tests, there is also evidence that cognitive ability tests may result in adverse impact. Specifically, there is evidence of score differences among racial groups (Ployhart & Holtz, 2008; Pyburn, Ployhart, & Kravitz, 2008; Roth, Bevier, Bobko, Switzer, & Tyler, 2001). While the general consensus is that these differences do not result in test bias, (see Schmitt, 2014), some authors have provided evidence for differential prediction (Aguinis, Culpepper, & Pierce, 2010; Berry, Clark, & McClure, 2011; Mattern & Patterson, 2013; te Nijenhuis & ven der Flier, 2000).

Another concern regarding the use of cognitive ability tests is the fact that they are susceptible to practice effects. When people take the same test more than

once, their scores improve by approximately one-fourth of a standard deviation (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007). While this change is not necessarily problematic (i.e., the change could be the result in a reduction of construct irrelevant variance such as test anxiety; Lievens, Reeve, & Heggestad, 2007), score changes do raise questions about test validity, namely, regarding whether initial or subsequent administrations are the best predictors of performance. If the initial administration is best, an implication might be that people should not be allowed to retest. If the second administration is best, an implication might be that everyone should be asked to take a practice test.

Acknowledging the fact that cognitive ability tests are not perfect and to improve decision-making ability, industrial–organizational psychologists have examined additional constructs that predict job performance. A broad construct that has been identified is personality. Research regarding the Five Factor model, arguably the most ubiquitous personality model, has consistently shown that personality traits, particularly conscientiousness and extraversion, are related to job performance (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). In addition to being valid on their own, personality traits offer incremental validity over and above tests of cognitive ability. Furthermore, personality offers another benefit: There is little evidence of racial subgroup score differences on personality measures, meaning that personality tests are less susceptible to adverse impact (Foldes, Duehr, & Ones, 2008; Hough, Oswald, & Ployhart, 2001).

The relative benefit of personality as a predictor of performance does not mean that the construct is perfect. A major criticism is that the validity

coefficients for each of the Big Five personality traits are low (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007b). There is also evidence that the coefficients are inconsistent from one study to the next (Barrick & Mount, 1993; Salgado, 1997), suggesting that there are moderators of the personality–performance relationship. Related to this, there is some emerging evidence that the relationship between personality and performance is curvilinear (LaHuis, Martin, & Avis, 2005; Le, Oh, Robbins, Iles, Holland, & Westrick, 2011), though these findings are inconsistent (Robie & Ryan, 1999). To further complicate matters, some researchers suggest that personality traits are not stable across the adult lifespan (Lucas & Donnellan, 2011; Roberts & DelVecchio, 2000; Wortman, Lucas, & Donnellan, 2012).

Authors have examined various methods for addressing low personality–performance validity coefficients. One area of research has focused on the granularity of the personality constructs that are used to predict performance. Early validity research explored broad personality traits. Over time, researchers have continued to refine personality models. One refinement has been the identification of narrow facets that underlie broad traits. To date, there is sufficient evidence that these facets can, in some instances, better predict performance than can higher-level traits (Hurtz & Donovan, 2000; Vinchur, Schippmann, Switzer, & Roth, 1998; Warr, Bartram, & Martin, 2005), especially when the facets are selected based on strong theory and a high-quality job analysis.

Situational strength has been examined as one of the moderators of the personality–performance relationship. Briefly, situational strength can be defined as the degree to which situations enhance or prevent the expression of behaviors consistent with one's personality (Meyer, Dalal, & Hermida, 2010; Mischel, 1977). This concept is based on the well-known perspective of interactionism, or the belief that the person and the environment jointly influence behavior (Tett & Burnett, 2003). Authors who have studied situational strength have, indeed, found that strong situations attenuate the impact of personality on behavior (Beaty, Cleveland, & Murphy, 2001; Cooper & Withey, 2009, Meyer et al., 2014). Conversely, in weak situations, in which behavior is less constrained by the environment, personality is more highly correlated with performance (Meyer, Dalal, & Bonaccio, 2009). Based on this body of evidence, it is possible to conclude that situational strength is an important moderator to consider when examining the validity of personality on performance. Thus, this project will further develop the literature by investigating the moderating effect of situational strength within a retest context.

Another major issue regarding the use of personality is that people are able to fake their scores on personality measures (Ones & Viswesvaran, 1998; Viswesvaran & Ones, 1999). This is true both in lab studies in which participants are asked to fake their scores (Viswesvaran & Ones, 1999) and in field studies that examine the faking behavior of job applicants (Birkeland, Manson, Kisamore, Barrick, & Smith, 2006).

Given the amount of research, across multiple research designs (e.g., between- and within-subjects designs, lab studies, field studies using incumbents and applicants), that documents the prevalence of faking, it is important to examine the impact that response distortion has on validity. Most of the work in this area suggests that faking does not have a significantly deleterious impact on validity coefficients (Barrick & Mount, 1996; Christiansen, Goffin, Johnston, & Rothstein, 1994; Hough, 1998; Ones & Viswesvaran, 1998); however, it is important to note that there are few studies that have examined this using a true retesting paradigm. The current project will help address this gap in the literature.

To summarize, this project has specific objectives. First, this work will contribute to the literature by further examining the validity of personality, operationalized at the facet level, within a retesting context. Second, this work will investigate this topic while also looking at the moderating impact of situational strength. Specifically, I will examine the degree to which situational strength has an impact on the magnitude of the change in scores across two administrations of a personality measure, as well as the effect that these changes may have on the validity of personality as a predictor of performance. Theoretically, this paper will address gaps in the literature identified above. Practically, the results of this project will help organizations determine the weight that personality should be given in selection systems. Significant results, for example, will further confirm that less weight should be given to personality when jobs could be classified as existing within strong situations. In addition, the results can help organizations determine their retesting policies. Specifically,

significant results will help confirm that allowing job applicants to retake personality measures does not undermine the utility and validity of existing selection practices.

Importance of Employee Selection

Employee selection is a vital process within organizations, as there are a number of positive outcomes associated with effective selection. Most importantly, high-quality selection processes enable organizations to recruit, select, and promote people who have the knowledge, skills, abilities, and other characteristics necessary to contribute to organizational success. In addition, selection is vital given the costs associated with human capital. Today, companies make a significant investment—in terms of salary, benefits, and training—in employees. A justification for this investment can be found in organizational theory. Human resource capital serves as an important source of competitive advantage across all levels (i.e., individual, department) of an organization (Barney & Wright, 1998; Nyberg, Moliterno, Hale, & Lepak, 2012; Ployhart & Moliterno, 2011).

Just as there are positive outcomes associated with effective selection, there are negative outcomes associated with ineffective selection. For example, when poor selection systems are used, the applicants who are hired may exhibit low person–job and person–organization fit. In addition, as a direct result of bad selection decisions and as an indirect result of poor fit, organizations are likely to experience increased turnover (Chatman, 1991). Most important, however, ineffective selection results in organizations hiring people who are unable to

contribute effectively and efficiently to organizational performance. Similarly, organizations may reject applicants who do have strong job-relevant knowledge, skills, abilities, and other characteristics. As a result of these positive and negative outcomes, it is important for organizations to use high-quality selection systems. Characteristics of such systems is that they are high in utility and use predictors that are both reliable and valid.

Multidimensionality of Job Performance

The belief that job performance is multidimensional is not new (Schmitt, 2014). In a well-known article, Austin and Villanova (1992) discussed what they referred to as "the criterion problem" (p. 836), which arises, in part, due to the fact that different criteria are used for different reasons. Despite the fact that operationalizing and measuring multidimensional constructs is difficult, however, does not mean that efforts should not be made to do so (Austin & Villanova, 1992). Other authors have also enumerated reasons for considering the dimensionality of criteria. Murphy and Shiarella (1997), for example, have identified three reasons: First, job performance is, by nature, multidimensional; second, different predictors help explain unique variance in performance; and third, there are different antecedents to different dimensions of performance.

Historically, a common approach to account for performance dimensionality has been to focus on both task performance and contextual performance (Borman & Motowidlo, 1997). Contextual performance also aligns with the concept of organizational citizenship behaviors (Borman & Motowidlo, 1997), another variable that has been used in validation research (Chiaburu, Oh,

Berry, Li, & Gardner, 2011; Organ & Ryan, 1995). Authors using the task performance–contextual performance approach have found that some predictors are more highly correlated with contextual performance than task performance (Hattrup, O'Connell, & Wingate, 1998).

Other approaches have also been used to account for the dimensionality of job performance. For example, one team of authors used a deductive approach to develop a specific taxonomy of managerial competencies (Tett, Guterman, Blier, & Murphy, 2000). Using existing taxonomies and feedback from subject matter experts, Tett et al. (2000) established a model comprised of 53 competencies organized into nine broad dimensions. In their manuscript, the authors pointed out that multidimensional models are useful because they can help practitioners better identify predictor constructs (Tett et al., 2000). Thus, the criterion that will be used in this research will assess multiple dimensions of performance.

Validity of Cognitive Ability Tests

Without a doubt, general mental ability is one of the best predictors of job performance, regardless of job type or industry (Hunter & Hunter, 1984; Schmidt, 2002; Schmidt & Hunter, 1998, 2004). Despite consistently high validity coefficients, however, cognitive ability tests are not perfect. A major issue is that they can result in adverse impact as a result of racial and ethnic subgroup score differences (Ployhart & Holtz, 2008; Pyburn et al., 2008; Roth et al., 2001). Meta-analytic evidence examining the differences in mean scores suggests that on cognitive ability tests, the score difference between Blacks and Whites is approximately one standard deviation and the score difference between Whites

and Hispanics is approximately .72 *d* (Ployhart & Holtz, 2008; Roth et al., 2001).

Another team of authors, looking at differences in validity coefficients rather than

score differences, also found meta-analytic evidence of subgroup differences that

can result in differential prediction. Specifically, the authors reported that validity

coefficients were .09 higher for Whites compared to Blacks, .04 higher for Whites

compared to Hispanics, and .01 higher for Whites compared to Asians (Berry et

al., 2011).

This evidence suggests that subgroup differences and the resulting

differences in validity continues to be an issue. Various strategies for addressing

this have been suggested. One strategy is to supplement cognitive ability tests

with non-cognitive predictors (Ployhart & Holtz, 2008) such as personality

(Schmidt & Hunter, 1998). The use of personality as a predictor will be discussed

in greater detail later in this paper.

Cognitive ability tests and retest effects. Another issue to consider

when using cognitive ability tests to predict job performance is retest effects,

which can be defined as "test score changes after prior exposure to an identical

test or to an alternative form of [a] test under standardized conditions" (Lievens,

Buyse, & Sackett, 2005, p. 982). People may be exposed to the same or multiple

forms of a test for various reasons. For example, both the *Uniform Guidelines for*

Employee Selection Procedures (1978) and the *Standards for Educational and*

Psychological Testing (American Educational Research Association, American

Psychological Association, and National Council on Measurement in Education,

2014) suggest that people be allowed to retest due to the high-stakes decisions

that are made based on people's scores. Many organizations and educational institutions allow applicants to retest as an issue of fairness.

Regardless of whether organizations officially allow applicants to retest, given the limited number of commercially available tests (e.g., the Hogan inventories, the Wonderlic Personnel Test), along with current employment trends in which few people remain with a single organization for their entire careers, it is likely that people will encounter the same test, or at least multiple tests of the same construct, more than once (Reeve & Lam, 2007). Authors have suggested that between 2% and 10% of external applicants retest and that around 32% of internal candidates retest for promotions (Villado, Randall, & Zimmer, 2016).

Regarding retest effects, there is sufficient evidence from both single studies (for example, see Dunlop, Morrison, & Cordery, 2011; Matton, Vautier, & Raufaste, 2011; Raymond, Neustel, & Anderson, 2007) and meta-analyses that scores do increase from one administration of a test to the next. Hausknecht et al. (2007) reported a mean effect size of size of $d = .24$, suggesting that across the first and second administration of a cognitive ability test, scores increase by about one-fourth of a standard deviation. What is especially problematic, given the previous discussion of adverse impact, is the fact that there is evidence of subgroup differences in score improvement. In a field study using data from more than 2,000 applicants, Schleicer, Van Iddekinge, Morgeson, and Campion (2010) found that there was greater score improvement for Whites compared to Blacks and Hispanics, for women compared to men, and for people under the age of 40 compared to people over the age of 40.

In the literature regarding retest effects on cognitive ability tests, authors have posited various explanations for score changes. One model suggests that there are three reasons for score improvement (Lievens et al., 2005, 2007). The first reason is that score changes reflect a true change in people's standing on the construct. In other words, people's scores improve because they develop cognitive ability. While this may be true of score changes on some predictors (i.e., people could develop job knowledge as a result of experience), this explanation is unlikely given the stability of the construct.

The second possible explanation is that score changes reflect a reduction in construct-irrelevant variance (Lievens et al., 2007). For example, people may be less anxious during the second administration of a test (Anastasi, 1981). Additional examples include test takers' level of motivation and the degree to which they remember items they encountered during the initial administration (Randall & Villado, 2017). The third possible explanation is that score changes reflect an increase in skills that are unrelated to the construct; an example of this is testwiseness (Lievens et al., 2007).

Given the scope of this project, it is important to discuss the impact that retest effects have on the validity of cognitive ability tests. While a high-level summary is that score changes appear to have little impact on validity, the reality is more nuanced. When analyzing data from a sample of students who had applied to medical school, Lievens et al. (2005) found that applicants who took a cognitive ability test a second time (i.e., who retested due to failing the initial administration of the test) did better on the second test (*d*, corrected for

unreliability = .46). However, the score changes did not result in a significant

change in validity coefficients, suggesting that the initial test and the retest could

be equally predictive of performance (Lievens et al., 2005).

Villado et al. (2016) reported similar results. In a lab study in which the

authors used the Wonderlic Personnel Test (WPT) as the predictor and a

composite proofreading and mathematically oriented catalogue task as the

criterion, retesting effects did not have a significant impact on validity

coefficients, regardless of the fact that there were score gains on the WPT. The

authors suggested that these findings may have been a result of score gains that

were approximately the same for all test takers; however, the authors did not

examine this suggestion empirically.

Despite the findings reported above, other researchers have found that

score changes associated with retest effects do have an impact on validity.

Specifically, there is evidence that retest scores on a job knowledge test,

compared to initial test scores, can be more predictive of performance for internal

employees applying for a promotion (Van Iddekinge, Morgeson, Schliecher, &

Campion, 2011). While one might assume that these results reflect an increase in

people's job knowledge as a result of experience, the authors of the study also

found that a sizable group of people who retook the job knowledge test did worse

on the second administration than they did on the first.

To summarize, cognitive ability is a predictor of job performance, but that

does not mean that tests of cognitive ability are not without issue in selection

contexts. A potential concern, given the scope of this project, is the impact of

retest effects on validity. However, the evidence suggests that retesting is only an issue in limited circumstances. In the next section of this paper, I will explore personality as a predictor of job performance.

Personality as a Predictor of Job Performance

Industrial–organizational psychologists working in the area of selection have identified performance predictors other than cognitive ability. Personality has received a significant amount of attention. The historical use of this construct, however, is varied. For most of the 20th Century, personality was ignored within applied psychology. Many current authors attribute this to the work of Ghiselli and Barthol (1953) and Guion and Gottier (1965). Ghiselli and Barthol (1953) concluded that while personality was predictive of performance, the evidence at the time suggested that the construct was useful for jobs in which personality was not important theoretically and that it was not useful for jobs in which personality was important theoretically. Based on a review of approximately 100 studies, Guion and Gottier (1965) were more definitive: They concluded that there was little evidence to support the use of personality measures to predict job performance.

While applied psychologists may have ignored personality, psychologists in other fields continued to study the construct. One of the most significant developments was the emergence of various taxonomies for organizing personality traits (for a history of taxonomy development, see Hough & Ones, 2001; John, 1990; and Mount & Barrick, 1995). Though multiple taxonomies exist (e.g., the PEN model [Eysenck, 1990], a three-factor model [Tellegen,

1985], and a six-factor model [Ashton & Lee, 2007]), arguably the most

ubiquitous is the Five Factor Model (FFM), also known as the Big Five Model,

which includes the traits of openness to experience, conscientiousness,

extraversion, agreeableness, and emotional stability (McCrae & Costa, 1987;

Tupes & Christal, 1961). A significant body of research suggests that the model

is stable across gender, language, and culture (Benet-Martinez & John, 1998;

Borkenau & Ostendorf, 1990; Digman & Takemoto-Chock, 1981; Katigbak,

Church, & Akamine, 1996; McCrae & Costa, 1997).

The development of the FFM enabled industrial–organizational

psychologists to explore personality vis-à-vis selection in new ways. Indeed, by

the 1990s, there were enough studies of the personality–performance relationship

that researchers could conduct meta-analyses. In a frequently cited work from the

period, authors reported that the five traits were predictive of performance across

various occupational families (i.e., professional, police, manager, sales,

skilled/semiskilled), with rhos ranging from .04 for openness to experience to .22

for conscientiousness (Barrick & Mount, 1991). Likewise, when averaged across

multiple criteria (i.e., job proficiency, training proficiency, personnel data), the

rhos ranged from .04 for openness to experience to .22 for conscientiousness

(Barrick & Mount, 1991). Based on the overall results of their meta-analysis,

Barrick and Mount (1991) concluded that conscientiousness is consistently

predictive of performance across all job families and criteria, extraversion is

predictive of performance for jobs involving social interactions (i.e., managers,

sales), and extraversion and openness to experience are predictive of training proficiency.

The results of a meta-analysis based on data collected from European samples, rather than US samples, were consistent with Barrick and Mount's (1991) results. Specifically, conscientiousness was predictive across occupational families (i.e., professionals, police, managers, skilled/unskilled) and criteria (supervisor ratings, training performance, personnel data; Salgado, 1997). When averaged across occupations and criteria, the rhos ranged from .02 for agreeableness to .22 for conscientiousness (Salado, 1997). Regarding the overall results, Salgado (1997) concluded that conscientiousness is the most effective performance predictor because of its generalizability. Another early meta-analysis looking at personality and performance, but not using the Five Factor Model, also found evidence of significant validity coefficients (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990).

Tett et al. (1991) contemporaneously published another meta-analysis of the personality–performance literature; their approach was more nuanced than the approach used by Barrick and Mount (1991). Specifically, Tett et al. (1991) examined various moderators of the personality–performance relationship. One of the most significant findings of their study was that when looking at confirmatory studies (i.e., studies in which the personality–performance relationship was theoretically driven rather than empirically driven), the validity coefficients for personality were larger than had been reported previously. When correcting for unreliability in predictors and criteria, the mean validity

coefficients reported by the authors ranged from .18 for conscientiousness to .33 for agreeableness (Tett et al., 1991). While Tett et al. (1991) found evidence of a personality–performance relationship, it is worth noting that their results are slightly different from those reported by Barrick and Mount (1991).

Conscientiousness as a predictor. Based on the research discussed above, there is sufficient evidence that personality is a useful predictor of performance. There is a general consensus that of the Big Five traits, conscientiousness is the best; it generalizes across occupational families and criteria (Barrick & Mount, 1991; Hough & Ones, 2001). At the same time, there is evidence that the conscientiousness–performance relationship can be mediated by other variables and that conscientiousness can act as a moderator. Using a small ($N = 91$) sample of sales representatives, Barrick, Mount, and Strauss (1993) found that goal setting and goal commitment mediate the conscientiousness–performance relationship. In this study, people high on conscientiousness were more likely to set goals and, as a result, were more likely to exhibit better performance than people low in conscientiousness (Barrick et al., 1993).

When looking at the moderating effects of conscientiousness, Bakker, Demerouti, and ten Brummelhuis (2012) found that conscientiousness moderated the relationship between work engagement and both task and contextual performance. In a separate study, Demerouti (2006) found evidence that conscientiousness moderates the relationship between flow, or a state of absorption, work engagement, and intrinsic motivation (Csikszentmihalyi, 1990)

and both task performance and contextual performance. In addition, there is evidence of an interactive effect between conscientiousness and agreeableness. In a study that included seven different samples, Witt, Burke, Barrick, and Mount (2002) found that in five of the samples and among employees who were high in conscientiousness, performance was lower for employees who were also low in agreeableness as opposed to high in agreeableness. In the remaining two samples, there was no conscientiousness–agreeableness interaction. Collectively, this research underscores the importance of understanding conscientiousness as it relates to and predicts job performance.

When examining the relationship between conscientiousness and performance, a limited number of criteria have been used. Most often, performance is operationalized as task performance (e.g., supervisor ratings, sales), contextual performance, and personnel data (e.g., absenteeism, turnover). As discussed above, when assessing test validity, it is important to consider the fact that job performance is multidimensional (Austin & Villanova, 1992). However, there are few studies in which authors have examined the validity of conscientiousness in relation to the dimensions of managerial performance that have been identified in the literature (Tett et al., 2000).

One of the few studies to have done this examined the relationships between conscientiousness, openness, and decision making. In the study, participants engaged in a series of decision-making tasks as part of a computer simulation. During the simulation, the task context changed such that the rules that were used to determine whether participants decisions were correct changed.

When analyzing the data, the authors found that participants who were low in conscientiousness actually made worse decisions after the rules changed (LePine, Colquitt, & Erez, 2000). Surprised by these results, the authors decided to further analyze the data by looking at the facets of conscientiousness. They found that post-rule-change performance was worse for people who scored high on order, dutifulness, and deliberation; the facets of achievement and self-discipline were not related to performance (LePine et al., 2000).

In another study, the authors operationalized the criterion by developing a measure to assess eight dimensions of managerial performance (i.e., planning, administration, development, communication, coordination, effort, organizational commitment, know-how; Barrick & Mount, 1993). The validity coefficients for conscientiousness were .32 and .25 for performance overall and performance by dimension, respectively (Barrick & Mount, 1993). The goal of the study was to examine the moderating impact of autonomy in the personality–performance relationship; thus, the authors did not provide the coefficients for each performance dimensions, nor did they explain how they calculated the validity coefficients for each performance dimension.

Finally, one study suggests that there is a negative relationship between conscientiousness and managerial competencies that include action, motivation, creativity, and communication, whereas there is a positive relationship between conscientiousness and managerial competencies that include organization, leadership, and analysis (Robertson, Gibbins, Baron, MacIver, & Nyfield, 1999). Collectively, the limited results of these studies suggest that additional research is

needed to further identify the predictive power of personality, particularly

conscientiousness, in relation to dimensions of managerial performance.

Furthermore, as has been pointed out, it is useful to theoretically link predictors

and criterion when conducting validation studies (Barrick & Mount, 1991).

However, good theory does not materialize from nowhere; empirical data can help

shape theory.

Measuring personality. When examining the validity of personality, it is

important to consider how the construct is operationalized. In the studies

discussed above, researchers frequently used general personality measures such as

the NEO (McCrae & Costa, 1987). There is some evidence that this is acceptable.

In one study, using measures designed specifically to measure the FFM traits, as

opposed to general personality measures, resulted in larger validity coefficients

for conscientiousness and emotional stability (Salgado, 2003). At the same time,

Hurtz and Donovan (2000) suggested that a methodological weakness of the

Barrick and Mount (1991) and Tett et al. (1991) meta-analyses was that the

research teams included studies in which personality was measured by using

scales that were not designed with the FFM in mind. Hurtz and Donovan (2000)

addressed this concern by conducting another meta-analysis. Their results were,

however, consistent of those of Barrick and Mount (1991) and Tett et al. (1991).

Specifically, the overall validity coefficients ranged from .06 for openness to .22

for conscientiousness (Hurtz & Donovan, 2000). The results were also similar to

the previous meta-analyses in regards to occupational families and criteria (Hurtz

& Donovan, 2000).

In addition to using measures specifically designed to assess the Big Five traits, researchers have examined the impact of contextualizing items. Doing so can result in better validity coefficients (Schmit, Ryan, Stierwalt, & Powell, 1995). In one study, the use of general items, as opposed to the use of items with a work-related frame of reference, resulted in a greater degree of measurement error (Schmit et al., 1995). Similarly, another team of researchers have found evidence to support the notion that people are more likely to use non-work frames of reference when responding to personality items that are not contextualized (Fisher, Cunningham, Kerr, & Allscheid, 2017).

Personality and adverse impact. As has been pointed out previously, one of the issues associated with the use of cognitive ability tests is that they are susceptible to racial and ethnic subgroup score differences, which can result in adverse impact (Ployhart & Holtz, 2008; Pyburn et al., 2008; Roth et al., 2001; Sackett, Schmitt, Ellingson, & Kabin, 2001). Researchers have investigated various ways to address this. One avenue of research has explored subgroup differences on personality measures (Oswald & Hough, 2011).

In a comprehensive review of the literature, Hough et al. (2001) reported that there were minimal racial subgroup differences at the trait level. The largest difference among racial groups included in the review—Blacks, Hispanics, American Indians, Asians, and Whites—was a $d = -.20$ difference between Blacks and Whites on openness, in which Blacks scored lower than Whites (Hough et al., 2001). Even at the facet level, the authors found minimal differences among racial groups (Hough et al., 2001).

In their review, Hough et al. (2001) also looked at gender and age differences. In general, the differences between women and men were small, with one exception: The difference between women and men on agreeableness was moderate at $d = .39$, with women scoring higher than men (Hough et al., 2001). The trait-level effects for age, using 40 as the cutoff, were also small, with two exceptions: Older adults' scores were .49 d, or about one-half a standard deviation, higher than younger adults scores for the dependability and .24 d lower for achievement (i.e., two facets of conscientiousness; Hough et al., 2001).

The results of a more recent meta-analysis provide a bit more insight into subgroup differences. Consistent with Hough et al. (2001), the trait-level differences between Blacks and Whites were small, with the largest difference being $d = -.16$ for extraversion, in which Blacks scored lower than Whites (Foldes et al., 2008). The trait-level differences between Asians and Whites were also small, with one exception: Asians scored considerably higher on agreeableness (Foldes et al., 2008). Similarly, the trait-level differences between Hispanics and Whites were minimal (Foldes et al., 2008). The trait-level differences between American Indians and Whites, however, were moderate, ranging from $d = -.21$ for emotional stability to $d = -.33$ for extraversion, with Whites scoring higher for all traits except for conscientiousness (Foldes et al., 2008). A caveat for this study is that the sample size for American Indians ranged between 70 for agreeableness and 743 for emotional stability (Foldes et al., 2008). Overall, the authors concluded that at the trait level, racial subgroup differences are minimal and are unlikely to result in adverse impact. At the same time, the authors remind us that

this likelihood also depends on the make-up of the applicant pool and the severity of the selection ratio (Foldes et al., 2008).

Taken together, the results of these studies suggest that using personality as part of a composite predictor can reduce the possibility of adverse impact (Barrick & Mount, 2005). The use of personality can also result in incremental validity, which will be discussed in the next section of this paper.

Personality and incremental validity. One way to increase the utility and usefulness of selection procedures is to include additional predictors that help explain unique variance in the criterion (Sechrest, 1963). In a summary of existing research, Schmidt and Hunter (1998) suggested that using integrity tests in addition to a test of cognitive ability can increase validity by as much as 27%. Likewise, including work samples and structured interviews can each increase validity by up to 24%, and adding a measure of conscientiousness can increase validity by 18% (Schmidt & Hunter, 1998).

Research focusing specifically on personality has, indeed, demonstrated the incremental validity of the construct (Barrick & Mount, 2005; Oswald & Hough, 2011). Among a small sample of incumbent accountants, Day and Silverman (1989) found that a measure of personality offered incremental validity over and above a test of cognitive abilities when predicting a composite measure of performance. This was true for personality as a global composite, for ascendancy (i.e., a proxy of cooperative orientation) as an incremental predictor of the incumbents' self-report ratings of potential success, for work orientation as an incremental predictor of client relations, and for interpersonal orientation as an

incremental predictor of cooperation (Day & Silverman, 1989). In a study examining the incremental validity of personality over biodata forms among a sample of insurance sales representatives, McManus and Kelly (1999) found that a measure of the Big Five resulted in an R^2 change of .08 for sales performance and an R^2 change of .16 for contextual performance.

When looking specifically at cognitive ability and conscientiousness, there is evidence that the personality trait offers incremental validity for a variety of criteria. In a study using archival employee data, Avis, Kudisch, and Fortunato (2002) found that using a measure of conscientiousness in addition to a measure of cognitive ability resulted in a change in R^2 of .08 when overall performance was used as the criterion. There was a .03 change in R^2 when task performance was used as the criterion, a .09 change in R^2 when contextual performance was used as the criterion, and a .04 change in R^2 when customer service was used as the criterion (Avis et al., 2002).

Mount, Barrick, and Strauss (1999) have also looked at conscientiousness as a source of incremental validity. Among a sample of US Army managers, including conscientiousness as a predictor, in addition to cognitive ability, resulted in an R^2 change of .06; among a sample of sales representatives, the change in R^2 was .07; and among a sample of managers, the change in R^2 was .06 (Mount et al., 1999). The authors also noted that there was not a significant conscientiousness–cognitive skills interaction, providing further evidence of the unique predictive power of conscientiousness (Mount et al., 1999). Rothstein and Goffin (2006) have pointed out that there are few studies, relatively speaking,

examining personality's incremental validity; they note that this is an area where additional research is needed.

Collectively, the studies discussed in this section of the paper provide evidence supporting the notion that personality is a valid predictor of performance. Specifically, using personality as part of a composite predictor can help address concerns regarding the potential for adverse impact that is associated with measures of cognitive ability. In addition, personality offers incremental validity over and above measures of cognitive ability. This is not to say that there are no issues associated with the use of the construct. Major concerns regarding the use of personality as a predictor will be discussed in the following section.

Addressing Limitations of Personality

As is true of cognitive ability, there are also limitations associated with the use of personality as a predictor of performance. A common critique of personality is that the validity coefficients reported in the literature are small to moderate (Murphy & Dzieweczynski, 2005; Schmidt & Hunter, 1998).

Facets versus traits. One way to address the concern regarding validity coefficients is to look at the validity of personality facets as opposed to broad traits (Oswald & Hough, 2011; Schmitt, 2014). Warr et al. (2005) conducted a study to examine this topic. Specifically, they explored the validity of conscientiousness and two of its facets (i.e., achievement orientation, dependability) and extraversion and two of its facets (i.e., potency, affiliation); they used sales as a criterion. Overall, the correlation between conscientiousness and sales was .23. However, the correlation between the achievement orientation

facet and sales was .27, while the correlation between the dependability facet and sales was only .05 (Warr et al., 2005). Likewise, the overall correlation between extraversion and sales was .08; the correlations for the potency facet was .20, and the correlation for the affiliation facet was -.05 (Warr et al., 2005). These results are similar to those reported in an earlier meta-analysis. When looking at the validity of personality for sales jobs, achievement and potency were better predictors than their respective broader traits, both when supervisor ratings and sales were used as criteria (Vinchur et al., 1998).

In addition to offering direct validity, there is meta-analytic evidence that personality facets also offer incremental validity over broad traits for some criteria. While facets of conscientiousness did not offer incremental validity for task performance, a team of authors found that the facets offered incremental validity when predicting components of contextual performance, such as job dedication, counterproductive work behaviors, and interpersonal facilitation (Dudley, Arvis, Lebiecki, & Cortina, 2006). In regards to occupational families, there is evidence that facets of conscientiousness offer incremental validity over conscientiousness measured at the trait level for sales, managers, and skilled/semiskilled employees (Dudley et al., 2006).

In a lab study in which organizational citizenship behaviors were used as criteria, one team of researchers reported that facets of extraversion, namely positive emotion and surgency, were better predictors than the trait (Moon, Hollenbeck, Marinova, & Humphrey, 2008). These findings, that facets can have

as much or more validity than traits, seems to hold true for other personality taxonomies as well (Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995).

Despite the evidence discussed above, not everyone agrees that facets are better predictors than traits. Based on a practical perspective (i.e., organizations have to somehow combine predictor scores to make real-world hiring decisions) and a theoretical perspective (i.e., when personality is operationalized at too narrow a level, it is unlikely the narrow operationalizes will generalize across jobs; it would be too labor intensive to theoretically link narrow constructs with performance criteria), Ones and Viswesvaran (1996) argue against the use of facets and for the use of traits. Empirically, there is at least one study in which three facets of conscientiousness (i.e., order, industrious, self-control) were not better predictors of performance than the broader traits (Salgado, Moscoso, & Berges, 2013).

Based on the literature as a whole, one can safely assume, with a few exceptions, that personality facets, rather than broad personality traits, are better at predicting performance. Authors have suggested that this is because trait-level measures are too heterogeneous; for example, low or negative facet-performance relationships can obscure important, valid positive facet–performance relationships (Hough, 1992; Judge, Rodell, Klinger, Simon, & Crawford, 2013; Tett, Steele, & Beauregard, 2003). In addition, the use of facets, rather than traits, allows for the development of more direct predictor–criterion linkages (Stewart, 1999).

Situational strength. Authors have identified another way to address the low validities associated with personality, namely by considering possible moderators of the personality–performance relationship. A possible moderator that has been suggested is situational strength (Barrick & Mount, 2005; Cooper & Withey, 2009).

The concept of situational strength is based on an interactionist perspective (Tett & Burnett, 2003), which is the belief that behaviors are the result of both individual differences and the environment (Hattrup & Jackson, 1996; Meyer et al., 2010; Mischel, 1977; Weiss & Adler, 1984). Specifically, situational strength refers to "the degree to which situational constraints are present in the environment" (Judge & Zapata, 2016, p. 1150). In strong situations, there are clear cues that provide information about what behaviors are expected. In other words, strong situations impede the expression of behaviors that are consistent with one's personality (Cooper & Withey, 2009). A classic example of a strong situation is waiting in traffic at a red light (Mischel, 1977). Within an organization, an example of a strong situation is an assembly line. Conversely, weak situations are those in which there is little guidance regarding the behaviors that are expected, thereby allowing for greater expression of behaviors that are consistent with one's personality (Cooper & Withey, 2009; Meyer et al., 2010). An example of a weak situation within an organization is a skunkworks project, in which an ad hoc team is tasked with creating innovative products or solutions to problems. These types of projects are generally not part

of an organization's larger research and design function, and they provide team members with a great deal of latitude in their approach to problem solving.

To help offer clarity regarding how the concept of situational strength applies to organizations, Meyer et al. (2010) developed a model consisting of four facets of situational strength. The first facet is clarity, which refers to "the extent to which cues regarding work-related responsibilities or requirements are available and easy to understand" (p. 125). Clarity can be provided in a number of ways, including the presence of unambiguous standardized operating procedures, organizational norms as expressed via culture, and task assignments and structure as provided by supervisors (Meyer et al., 2010). The second facet is consistency, or the degree to which cues about work responsibilities are compatible with one another. Contextual cues can provide either similar or dissimilar information about behaviors that are expected (Meyer et al., 2010). The third facet is constraints, or whether employees have discretion over their actions and decisions (Meyer et al., 2010). Finally, the fourth facet is consequences, or the degree to which an employee's actions will impact organizational stakeholders (Meyer et al., 2010). Thus, according to this model, strong situations are marked by high clarity, high consistency, high constraints, and high consequences.

In relation to personality and job performance, situational strength should theoretically act as a moderator (Judge & Zapata, 2016). There are a handful of studies that have examined this empirically. In an exploratory field study in which there was a small sample size ($N = 58$), there was initial evidence that the

correlations between the Big Five traits and a measure of contextual performance was lower when participants perceived their respective job performance situations to be strong as opposed to weak (Beaty et al., 2001). Other authors have found that the correlation between conscientiousness and organizational citizenship behaviors (OCB) and between agreeableness and OCB were lower when situational strength was perceived to be strong (Meyer et al., 2014). The same findings have been reported when a measure of counterproductive work behaviors was used as the criterion (Smithikari, 2008). Finally, Gellatly and Irving (2001) reported that the relationship between extraversion and contextual performance and between agreeableness and contextual performance was moderated by perceptions of autonomy, which is theoretically aligned with the constraints component of Meyer et al.'s (2010) model of situational strength.

There is also evidence that situational strength acts as a moderator of the personality–performance relationship when task performance is used as the criterion. In a study examining the impact of autonomy on the relationship among a sample of 154 people participating in a US Army training program, the criterion was operationalized by measuring eight dimensions of managerial performance (i.e., planning, administration, development, communication, coordination, effort, organizational commitment, know-how; Barrick & Mount, 1993). In the study, the authors found that conscientiousness, extraversion, and agreeableness had larger validity coefficients when autonomy was high as opposed to low (Barrick & Mount, 1993).

In a meta-analysis examining the impact that situational strength has on the conscientiousness–performance relationship, situational constraints acted as a moderator when overall performance was used as the criterion but not when task performance was used as the criterion (Meyer et al., 2009). In addition, the conscientiousness–performance relationship was stronger when employees' actions and decisions had little consequence, both when overall performance and task performance were used as a criterion (Meyer et al., 2009). A weakness of the Meyer et al. (2009) meta-analyses is that the authors examined personality at the trait level rather than at the facet level. In a more recent meta-analysis, Judge and Zapata (2016) reported that the relationship between all Big Five constructs and performance were stronger when situations were weak (i.e., operationalized as the process by how work was done); the relationships between agreeableness and performance and between openness and performance were also stronger when situations were weak (i.e., operationalized as the outcomes of work).

Personality and Response Distortion

When discussing the use of personality to predict job performance, we cannot ignore the issue of faking. In this context, faking can be defined as a "focused, intentional effort to respond in a socially desirable manner on a personality test in order to increase the likelihood of obtaining a job" (Berry & Sackett, 2009, p. 835). This has also been referred to as socially desirable responding, response distortion, and impression management (Griffith, Chmielowski, & Yoshita, 2007). Partially as a result of the faking concern, some people question whether personality should be used in selection or suggest that it

should only be used with extreme caution (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007a, 2007b).

In general, there is sufficient evidence that people can fake personality measures (Hough & Ones, 2001; Ones & Viswesvaran, 1998; Viswesvaran & Ones, 1999). At the same time, there is evidence that these effects can be minimized via the use of various interventions, such as the use of forced-choice measures (Jackson, Wroblewski, & Ashton, 2000) and warnings (Adair, 2014).

Faking research is generally done using an "instructed to fake" paradigm in which participants are asked to complete a personality inventory honestly and are then asked to fake good (i.e., to present themselves in the best possible light) and/or fake bad (i.e., present themselves in the worse possible light). The size of the differences between the honest scores and the faked scores provide information about the magnitude by which people can fake their scores when they are motivated to do so. Another way to assess the impact of faking is to measure social desirability directly (Ones & Viswesvaran, 1998). Some measures, such as the Minnesota Multiphasic Personality Inventory, have subscales that measure this construct.

A meta-analysis of studies that have used the "instructed to fake" paradigm clearly indicate that faking is possible, with the effects being larger for within-subjects designs as opposed to between-subjects designs (Viswesvaran & Ones, 1999). Despite the evidence that people can fake, the authors of the meta-analysis point out that the data do not mean that people do fake in the real world. This point has been made by other authors as well (Griffith et al., 2007).

Another research paradigm that has been used to assess faking, and that is especially relevant to selection contexts, is to compare applicants' personality scores with incumbents' scores. Data from this approach suggest that applicant scores are, indeed, higher than incumbent scores for the traits of emotional stability, extraversion, conscientiousness, and openness (Birkeland et al., 2006). There is also evidence that effect sizes are larger for traits that employees perceive to be job relevant. For example, the effect for extraversion is larger for sales people (Birkeland et al., 2006). The effects associated with this paradigm are smaller than the faking effects associated with the "instructed to fake" paradigm.

Finally, retesting is a paradigm through which to explore score changes on personality measures. There are a number of studies that have used this approach. The evidence is clear: As with cognitive ability tests, there is evidence of retest effects for personality measures (Hausknecht, 2010). A meta-analysis reported significant mean differences for agreeableness, extraversion, conscientiousness, and openness to experience (Halpert, Gerjerts, Miller, Lukasik, & Fritts, 2008). It is important to know that while there are score changes, they are not always in a positive direction (Halpert et al., 2008; Hogan, Barrett, & Hogan, 2007). This may be the result of people not knowing that organizations want in terms of a "best score" (Halpert et al., 2008)

Research using this paradigm has provided other interesting information. First, given the fact that retest effects exist, it is not surprising that people who fail the first administration of a personality measure engage in alternative response strategies on subsequent administrations (Hausknecht, 2010; Landers, Sackett, &

Tuzinski, 2011). Second, internal, as opposed to external, candidates

(Hausknecht, 2010) and people high in conscientiousness (Barron, Randall, Trent,

Johnson, & Villado, 2017) are more likely to retest when given the chance to do

so. Third, people who receive direct negative feedback (i.e., those who are told

that their personality scores are the reason they did not receive a job offer) are

more likely to exhibit larger score differences (Holladay, David, & Johnson,

2013). However, there is also evidence that retest scores for people who are

asked to retest are more accurate, or closer to a baseline measure administered

outside a selection context (Ellingson, Heggestad, & Makarius, 2012). Finally,

data from a simulation study indicate that the weight given to personality (i.e., in

relation to other predictors) and the selection ratio that is used have an impact on

the size of the score differences between failing and passing applicants (Walmsley

& Sackett, 2013).

Response distortion and validity. Faking is potentially an issue in

selection contexts because it could have an impact on validity. One approach to

assess this is to administer measures of personality and social desirability. Then,

participants' personality scores are either adjusted based on the social desirability

scores or they are removed from the applicant pool. The results of the studies

using this approach are mixed. When using conscientiousness and work

orientation as predictors and effort and leadership as criteria, the validity

coefficients are larger for honest responders (Hough et al., 1990). In other

studies, there are minor, inconsequential differences in validity coefficients

(Barrick & Mount, 1996; Christiansen et al., 1994; Hough, 1998; Ones & Viswesvaran, 1998).

Faking could also be an issue to the degree that it alters the rank order of applicants (Berry & Sackett, 2009). Some authors have reported that faking does have an impact on rank order (Krammer, Sommer, & Arendasy, 2017) and that this impact is magnified when more stringent selection ratios are used (Griffith et al., 2007). In addition, Hogan et al. (2007) reported that the correlation between initial administrations of a personality measure and subsequent administrations ranged from .46 to .68.

Rationale

It is clear that under certain conditions, personality can be a useful predictor of job performance: The construct is less prone to sub-group differences, and it offers incremental validity over and above other predictors. In addition, there is some evidence that the use of more narrow facets can mitigate the issues associated with using broader traits. By examining the facets of conscientiousness in relation to a criterion, this study will continue to develop this literature.

Researchers have suggested that situational strength may moderate the personality–performance relationship, such that the relationship is stronger in weak situations. Most previous studies have examined this issue at the trait level. Thus, examining the personality–performance relationship at the facet level, while also experimentally manipulating situational strength will help us better understand the moderating role of this construct.

Finally, there has been a great deal of research examining retesting effects related to both cognitive skills tests and personality tests. Most of the research examining personality tests has used an "instructed to fake" paradigm in which research participants were explicitly instructed to provide inaccurate scores. This study will help develop the literature by examining the impact that retesting has on validity. One specific benefit of this study is that personality will be measured at the facet level. Another benefit is that multiple criteria will be used.

Statement of Hypotheses and Research Questions

Hypotheses 1–VI: There will be a main effect for facet-level conscientiousness, such that the six facets of conscientiousness (i.e., competence, order, dutifulness, achievement striving, self-discipline, cautiousness; Costa, McCrae, & Dye, 1991) will better predict performance on the in-basket criteria (i.e., coordination, decisiveness, information management, problem awareness) than will trait-level conscientiousness. This will be true for both the initial test scores (i.e., Hypotheses Ia–VIa) and the retest scores (i.e., Hypotheses Ib–VIb).

Hypotheses VII–XII: Situational strength will moderate the personality–performance relationship. This will be true for both the initial test scores (i.e., Hypotheses VIIa–XIIa) and the retest scores (i.e., Hypotheses VIIb–XIIb) across all four in-basket performance criteria.

Hypotheses XIII–XVIII: The facet-level personality–performance correlations for the initial personality measure will not be significantly different from the facet-level personality–performance correlations for the retest. This will be true for all four in-basket performance criteria.

Research Questions I–VI: Is there evidence that the relationship between facet-level personality and performance on the in-basket criteria is nonlinear rather than linear?

Method

Participants

Participants were recruited via Amazon Mechanical Turk (mTurk), an online platform that can be used to post small tasks that can be done remotely. There is evidence that data collected from mTurk samples are more demographically diverse than tradition convenience samples (Barger, Behrend, Sharek, & Sinar, 2011; Landers & Behrend, 2015). In addition, data are generally reliable (Buhrmester, Kwang, & Gosling, 2011), and researchers have been able to replicate studies using traditional samples with data collected from mTurk (Paolacci, Chandler, and Ipeirotis, 2010). The participants were compensated $3.00 for completing the study. In order to participate, participants had to be over the age of 18, live in the United States, and be fluent in English.

Assuming alpha = .05, power = .80, and a small effect size (i.e., consistent with the effect sizes reported in the literature review and the conscientiousness test validity coefficient of .31 reported by Schmidt & Hunter [1998]), Cohen (1992) suggests a minimum of 783 participants when conducting correlations; he suggests that for other tests (e.g., mean difference, ANOVA and multiple regression) and assuming the same conditions, fewer participants are needed.

Data for the study were collected from 1,574 participants. However, data from 619 participants were excluded from the study; these participants were

excluded because at least one of the answers they provided to the attention- and manipulation-check items was incorrect, they completed the study more than once, or they did not complete the study fully (e.g., responses were provided for a single scale only). The mean age of the 955 participants who were retained was 37.48 years (SD = 10.61). Regarding sex, 51.52% of the participants identified as male, 48.18% identified as female, and 0.30% either did not report their sex (0.20%) or indicated that they preferred a term other than *male* or *female* (0.10%). Regarding ethnicity, 82.20% of the sample identified as Caucasian; 9.11% identified as African American; 6.18% identified as Hispanic, Latino, or Latina; 6.81% identified as Asian or Pacific Islander; 1.36% identified as Native American or Alaskan Native; and 0.63% identified as Other. Note that the percentages add up to more than 100.00%; participants were able to select more than one option. The sex and ethnicity breakdown by experimental condition is reported in Table 1.

Table 1

Demographic Breakdown by Situational Strength

Demographic Variables	Strong	Weak	Total
Male	257	235	492
Female	227	233	460
Other or did not respond	2	1	3
Caucasian	397	388	785
African American	47	40	87
Hispanic, Latino, Latina	35	24	59
Asian, Pacific Islander	30	35	65
Native American, Alaskan Native	7	6	13
Other or did not respond	3	3	6

Note. N = 955. Ethnicity adds up to more than 955; participants could select more than one option.

To get a sense of participants' work lives, they were asked to indicate the average number of hours they work per week. Of the sample, 8.60% indicated that they were unemployed, 4.00% indicated that they work less than 20 hours per week, 49.70% indicated working 21–40 hours per week, 39.50% indicated working 41–60 hours per week, 1.90% indicated working more than 60 hours per week, and 0.30% of the sample did not respond to this item. This information is summarized in Table 2.

Table 2

Participants' Average Number of Hours Worked Per Week

Number of Hours	N	Percent of Participants
Unemployed	82	8.60
< 20 hours	38	4.00
21–40 hours	475	49.70
41–60 hours	339	35.50
> 60 hours	18	1.90
Did not respond	3	0.30

Note. $N = 955$.

Materials

Facet-level conscientiousness scale. A 42-item scale developed using items from the International Personality Item Pool (IPIP) was used. All 6 facets of conscientiousness that are part of the NEO were assessed. Borrowing from the faking literature, a warning appealing to participants' moral principles (e.g., "Thank you for taking the time to fully read each item and to answer to the best of

your ability. Your efforts will help us better understand the role of personality within modern organizations.") was included in the directions. Consistent with Fisher et al. (2017) and Schmidt et al. (1995), the scale items were contextualized to provide a work-related frame of reference. Participants were asked to use a 5-point scale to rate their agreement with each item. The items were administered in a random order to help prevent participants from identifying the facets that were assessed. The same pool of items was used for the initial test and for the retest. This scale can be found in Appendix A.

Criteria. An in-basket task developed by Anseel and Lievens (2006) and adapted from Tett et al. (2004) was used as the criterion. This in-basket has been designed to assess four competencies of managerial performance, namely problem awareness, coordinating, information management, and decisiveness. Participants' responses were scored based on a rubric developed by Tett, Menard, Guterman, and Beauregard (2001).

When completing the in-basket, participants were asked to imagine that they were managers of a paint manufacturing company, and they were given background information about their role, along with a calendar and a company directory. Given the reading-intensive nature of this criterion, participants were asked to complete three attention-check items before they completed the criterion. To complete the criterion, participants read 10 emails, each of which included four response options.

For each email, each response is either a positive or a negative expression of one of the four performance dimensions (i.e., coordination, decisiveness,

information management, and problem awareness). The in-basket task is set up

such that each email assesses two of the four performance dimensions. Using the

first email as an example, the first response option is a negative expression of

decisiveness and was initially coded as a -1 for that dimension, the second

response option is a positive expression of problem awareness and was initially

coded as a +1 for that dimension, the third response option is a negative

expression of problem awareness and was initially coded as a -1 for that

dimension, and the fourth response option is a positive expression of decisiveness

and was initially coded as a +1 for that dimension.

Participants were asked to use a 5-point scale to rate how likely they were

to use each response option. To score the responses, participants' Likert-type

ratings were multiplied by the initial coding, described above, by using the

following calculations: *very unlikely* (multiply by -2), *unlikely* (multiply by -1),

likely nor unlikely (multiply by 0), *likely* (multiply by +1), *very likely* (multiply by

+2). Continuing to use the first email as an example, if a participant indicated that

he or she would be unlikely to use the first response option (i.e., selected *very*

unlikely, or 1), then the initial -1, representing a negative expression of

decisiveness, was multiplied by -2 to end up with a score of +2 for decisiveness.

The score for each performance dimension can be calculated by summing the

multiplied scores for the relevant response options across all the emails. For

example, to calculate a score for decisiveness, the multiplied responses for the

first and the fourth response options for email one, the first and the fourth

response options for email two, the first and the fourth response options for email

four, the third and the fourth response options for email five, and the first and the

second response options for email seven are summed.

Demographic questionnaire. So that the participant sample could be

described (i.e., these variables were not used to test the research hypotheses),

participants were asked to provide responses to demographic questions, including

gender, age, ethnicity, and employment status. This questionnaire can be found in

Appendix C.

Manipulation. Participants were assigned to one of two experimental

conditions in which situational strength was manipulated. In the first condition,

the strong condition, participants were led to believe that their work environment

will be highly constrained and that there is little room for the meaningful

expression of individual differences. In the second condition, the weak condition,

participants were led to believe that their work environment will not be

constrained and that expression of individual differences has an impact on

organizational outcomes.

The manipulation was based on the four facets of situational strength

identified by Meyer et al. (2010; i.e., clarity, consistency, constraints, and

consequences). All the participants were asked to imagine that they were

applying for a job in a field that interests them. Participants assigned to the

strong-situation condition, however, were told that clarity, for example, is

constrained by the environment; specifically, the manipulation told participants

that work-related responsibilities are clearly explained to all employees.

Participants assigned to the weak-situation condition were told that information about work-related responsibilities are not explained to employees clearly.

Participants were asked to complete a manipulation check. Specifically, participants were asked to provide responses to four items (i.e., one item per each facet of situational strength identified in Meyer et al.'s [2010] model) that best describe the organization that they read about in manipulation. Participants were excluded from the study if any of the items was answered incorrectly. The two manipulations and the manipulation-check items can be found in Appendix D.

Attention check. Given the amount of reading required of participants, the participants were asked to complete an attention check. Using recommendations provided by Oppenheimer, Meyvis, and Davidenko (2009), participants were shown a 5-item organizational attraction scale. In the directions, however, participants were asked to respond to each item by selecting *strongly disagree.* Participants were excluded from the study if any of the response options was incorrect. The attention check can be found in Appendix E.

Procedure

Participants were recruited via mTurk. After people signed up for the mTurk HIT, they were directed to a survey that was administered via Qualtrics. Recently, researchers have expressed concerns regarding the quality of MTurk data (see Hauser, Paolacci, & Chandler, in press). To help mitigate concerns regarding bots, at the end of the survey participants were assigned a unique, Qualtrics-generated code that they were required to enter. Participants were

informed of this in the Amazon HIT; only those participants who entered the code were compensated.

After reading the informed consent information, participants were provided with some basic information (i.e., participants were asked to imagine that they were applying for a new job) and were shown the experimental manipulation. Half the participants saw the strong situation, and half the participants saw the weak situation. After the participants read the information provided in the manipulation, they were asked to complete the personality measure described above, followed by the in-basket task. To help increase participants' motivation, they were told that they could earn a $0.25 "signing" bonus if they are found to be a good fit for the job to which they are applying. All participants, however, actually received the additional payment.

Next, the participants were told to imagine that they did not receive the job to which they had applied. They were also reminded that they should imagine that they needed a job because their current job was going to end in 4 weeks. Finally, they were told to imagine that they saw a similar job at the company to which they had just applied and that to be considered for this job, they needed to complete another personality scale. The personality measure was then re-administered to participants. To prevent memory effects, items were administered in a random order.

Finally, participants were asked to answer the demographic questionnaire. The participants were thanked for their participation, and they were told that their mTurk account would be credited within 7 days.

Results

The scale reliabilities, means, standard deviations, skewness, and kurtosis values are reported in Table 3. In general, the scales demonstrated acceptable reliability (i.e., Cronbach's $\alpha > 0.70$; Nunnally & Bernstein, 1994). An exception to this is the reliability coefficients for the four dimensions of the criterion (i.e., coordination, decisiveness, information management, problem awareness). The low reliability for the criterion is likely a result of the way in which the scale scores were calculated, as explained in Appendix B (i.e., each response option for each email was scored using a 5-point scale; the raw responses were multiplied by values ranging from -2 to +2, depending on the performance dimension assessed via each response; the scale scores were calculated so that the scores ranged from 1 to 20).

Normality was assessed using the guidelines recommended by Kline (2011); absolute values greater than 3.00 suggest that items might be skewed, and absolute values greater than 10.00 suggest that items may be kurtotic. Based on these criteria, none of the scale scores appear to be skewed, whereas the discipline retest score appeared to be slightly kurtotic. Because there was only a slight degree of kurtosis for a single variable, the data were not transformed. The correlation coefficients for the study variables are reported in Table 4.

Table 3

Reliability, Means, Standard Deviations, Skewness, and Kurtosis of Conscientiousness Facets and Criteria

Variable	α	*M*	*SD*	Skewness		Kurtosis	
				Statistic	SE	Statistic	SE
Competence	.83	1.36	.45	2.17	.08	8.30	.16
Competence Retest	.86	1.29	.47	2.34	.08	6.83	.16
Order	.84	1.75	.66	0.80	.08	0.03	.16
Order Retest	.85	1.60	.64	1.11	.08	0.83	.16
Duty	.77	1.32	.41	2.03	.08	5.21	.16
Duty Retest	.77	1.28	.41	2.30	.08	6.46	.16
Achievement	.84	1.29	.50	2.61	.08	8.39	.16

Variable	α	*M*	*SD*	Skewness		Kurtosis	
				Statistic	SE	Statistic	SE
Achievement Retest	.85	1.28	.49	2.67	.08	9.35	.16
Discipline	.87	1.33	.53	2.78	.08	9.64	.16
Discipline Retest	.87	1.28	.51	3.00	.08	10.51	.16
Caution	.78	1.41	.51	2.11	.08	5.56	.16
Caution Retest	.81	1.33	.50	2.58	.08	7.63	.16
Conscientiousness	.95	1.41	.41	1.97	.08	5.12	.16
Conscientiousness Retest	.95	1.34	.42	2.30	.08	6.51	.16
Coordination	.37	9.42	1.75	-0.20	.08	0.79	.16

Variable	α	*M*	*SD*	Skewness		Kurtosis	
				Statistic	SE	Statistic	SE
Decisiveness	.34	8.05	2.10	-.01	.08	.11	.16
Information Management	.43	5.64	2.37	.54	.08	.41	.16
Problem Awareness	.33	5.59	2.30	.40	.08	-.09	.16

Note. The range for *N* is 919–952, due to missing values. The personality measure used a 5-point scale.

Table 4

Correlations Among Study Variables

Variable	1	2	3	4	5	6	7	8
1. Competence	--							
2. Competence Retest	.73**	--						
3. Order	.40**	.36**	--					
4. Order Retest	.43**	.55**	.73**	--				
5. Duty	.67**	.59**	.49**	.48**	--			
6. Duty Retest	.64**	.73**	.39**	.58**	.75**	--		
7. Achievement	.78**	.73**	.39**	.44**	.71**	.68**	--	
8. Achievement Retest	.73**	.78**	.34**	.50**	.60**	.74**	.81**	--

Variable	9	10	11	12	13	14	15	16
1. Competence	.74**	.68**	.58**	.52**	.83**	.73**	.05	.22**
2. Competence Retest	.67**	.75**	.52**	.63**	.71**	.87**	.02	.32**
3. Order	.43**	.35**	.53**	.43**	.70**	.53**	-.03	.10**
4. Order Retest	.46**	.50**	.49**	.52**	.64**	.75**	.01	.16**
5. Duty	.70**	.62**	.73**	.59**	.85**	.71**	.04	.21**
6. Duty Retest	.64**	.73**	.61**	.71**	.73**	.87**	.06	.23**
7. Achievement	.80**	.75**	.63**	.57**	.85**	.77**	.07*	.21**
8. Achievement Retest	.71**	.80**	.54**	.63**	.74**	.87**	.06	.20**

Variable	17	18
1. Competence	.26**	.31**
2. Competence Retest	.27**	.32**
3. Order	.13**	.12**
4. Order Retest	.15**	.19**
5. Duty	.30**	.32**
6. Duty Retest	.31**	.32**
7. Achievement	.27**	.32**
8. Achievement Retest	.27**	.31**

Variable	1	2	3	4	5	6	7	8
9. Discipline	.74**	.67**	.43**	.46**	.70**	.64**	.80**	.71**
10. Discipline Retest	.68**	.75**	.35**	.50**	.62**	.73**	.75**	.80**
11. Caution	.58**	.52**	.53**	.49**	.73**	.61**	.63**	.54**
12. Caution Retest	.52**	.63**	.42**	.52**	.59**	.71**	.57**	.63**
13. Conscientiousness	.83**	.71**	.70**	.64**	.85**	.73**	.85**	.74**
14. Conscientiousness Retest	.73**	.87**	.53**	.75**	.71**	.87**	.77**	.87**
15. Coordination	.05	.02	-.03	.01	.04	.06	.07	.06
16. Decisiveness	.22**	.23**	.10**	.15**	.21**	.23**	.21**	.20**

Variable	9	10	11	12	13	14	15	16
9. Discipline	--							
10. Discipline Retest	.81**	--						
11. Caution	.69**	.57**	--					
12. Caution Retest	.57**	.67**	.70**	--				
13. Conscientiousness	.87**	.75**	.84**	.68**	--			
14. Conscientiousness Retest	.76**	.87**	.64**	.82**	.84**	--		
15. Coordination	.07*	.06	.05	.02	.04	.04	--	
16. Decisiveness	.21**	.22**	.21**	.22**	.23**	.24**	.13**	--

Variable	17	18
9. Discipline	.25**	.31**
10. Discipline Retest	.27**	.30**
11. Caution	.30**	.30**
12. Caution Retest	.31**	.32**
13. Conscientiousness	.28**	.32**
14. Conscientiousness Retest	.31**	.34**
15. Coordination	.05	.08*
16. Decisiveness	.22**	.32**

Variable	17	18
17. Information Management	--	
18. Problem Awareness	.32**	--

Note. N varies from 906 to 951 due to missing values. * = *p* < .01. ** = *p* < .001.

Hypotheses I–VI

The first set of hypotheses state that there will be a main effect for facet-level conscientiousness, such that facet-level scores, as opposed to trait-level scores, will better predict criterion performance. Rather than assessing the difference between two Pearson correlations, as was suggested in the proposal, these hypotheses were assessed by using Steiger's (1980) test for dependent correlations. This approach allows one to assess whether the difference between two dependent correlations (i.e., the correlation between facet-level conscientiousness and criterion performance and the correlation between trait-level conscientiousness and criterion performance) is statistically significant. During the analyses, the correlations are converted into z scores, and the difference between the scores is assessed.

The analyses were calculated by using an on-line calculator (Hoerger, n.d.) in which four data points were used for each analysis: the correlation between the facet-level scores and the criterion performance dimension, the correlation between the trait-level score and the criterion performance dimension, the correlation between the facet-level score and the trait-level score, and the sample size. Separate analyses were conducted for each facet of conscientiousness against each dimension of performance for both the initial personality test and the retest (i.e., 48 analyses were conducted).

The results of the analyses fell into three categories. First, the hypotheses were only supported for the discipline facet of conscientiousness when predicting coordination on the initial personality test. Second, the hypotheses were not

supported such that the trait-level score, rather than the facet-level score, better predicted (i.e., was more strongly correlated with) criterion performance for competence when predicting information management (personality retest); for order when predicting coordination (personality initial test), decisiveness (personality initial test and retest), information management (personality initial and retest), and problem awareness (personality initial and retest); for achievement striving when predicting decisiveness (personality retest), information management (personality retest), and problem awareness (personality initial and retest); and for discipline when predicting information management (personality retest) and problem awareness (personality retest). Third, the results were not statistically significant for the remaining relationships, suggesting that facet-level conscientiousness is neither better nor worse at predicting criterion performance than trait-level conscientiousness. Overall, the results suggest that facet-level scores are not better predictors than trait-level scores. The results are summarized in Table 5.

Table 5

Differences Between Facet–Criterion and Trait–Criterion Correlations

Facet–Performance Dimension	Z Score*	Significance
Competence–Coordination Pretest	0.83	0.41
Competence–Coordination Posttest	-1.42	0.16
Competence–Decisiveness Pretest	-0.42	0.67
Competence–Decisiveness Posttest	-1.10	0.27
Competence–Information Management Pretest	-1.13	0.26
Competence–Information Management Posttest	-2.48	**0.01**
Competence–Problem Awareness Pretest	-0.65	0.51
Competence–Problem Awareness Posttest	-1.19	0.23
Order–Coordination Pretest	-2.80	**0.05**
Order–Coordination Posttest	-1.10	0.23
Order–Decisiveness Pretest	-5.05	**0.01**
Order–Decisiveness Posttest	-3.88	**0.01**
Order–Information Management Pretest	-6.11	**0.01**
Order–Information Management Posttest	-6.90	**0.01**
Order–Problem Awareness Pretest	-8.35	**0.01**
Order–Problem Awareness Posttest	-6.68	**0.01**

Differences Between Facet–Criterion and Trait–Criterion Correlations

Facet–Performance Dimension	Z Score*	Significance
Duty–Coordination Pretest	0.00	1.00
Duty–Coordination Posttest	1.02	0.31
Duty–Decisiveness Pretest	-1.13	0.26
Duty–Decisiveness Posttest	-0.99	0.32
Duty–Information Management Pretest	1.04	.030
Duty–Information Management Posttest	-0.06	0.95
Duty–Problem Awareness Pretest	-0.70	0.49
Duty-Problem Awareness Posttest	-1.21	0.23
Achievement–Coordination Pretest	1.88	0.06
Achievement–Coordination Posttest	1.21	0.23
Achievement–Decisiveness Pretest	0.79	0.43
Achievement–Decisiveness Posttest	-2.60	**0.01**
Achievement–Information Management Pretest	-0.69	0.49
Achievement–Information Management Posttest	-2.77	**0.01**
Achievement–Problem Awareness Pretest	-11.64	**0.00**
Achievement-Problem Awareness Posttest	-9.42	**0.00**

Differences Between Facet–Criterion and Trait–Criterion Correlations

Facet–Performance Dimension	Z Score*	Significance
Discipline–Coordination Pretest	2.09	**0.04**
Discipline–Coordination Posttest	1.02	0.31
Discipline–Decisiveness Pretest	-1.16	0.25
Discipline–Decisiveness Posttest	-1.29	0.20
Discipline–Information Management Pretest	-1.92	0.05
Discipline–Information Management Posttest	-2.50	**0.01**
Discipline–Problem Awareness Pretest	-0.75	0.45
Discipline–Problem Awareness Posttest	-2.52	**0.01**
Caution–Coordination Pretest	0.91	0.36
Caution–Coordination Posttest	-1.18	0.24
Caution–Decisiveness Pretest	-0.66	0.51
Caution–Decisiveness Posttest	-1.06	0.29
Caution–Information Management Pretest	0.00	1.00
Caution–Information Management Posttest	0.22	0.83
Caution–Problem Awareness Pretest	-1.35	0.18
Caution–Problem Awareness Posttest	-1.36	0.18

Note. N varies from 906 to 926 due to missing values. *Positive scores indicate that facet-level conscientiousness, rather than trait-level conscientiousness, better predicts performance on the criterion; negative scores indicate the opposite. Significant correlations are set in bold type.

Hypothesis VII–XII

The second set of hypotheses state that the facet-level personality–criterion relationship is moderated by situational strength. These hypotheses were tested using hierarchical linear regression. In the first step, the facet-level personality variable and the dummy-coded situational strength variable were entered into the regression model. In the second step, the interaction term was entered into the model. Facet-level personality scores were centered around the mean before the product terms were created to reduce collinearity effects. A significant change in R^2 between the two models suggests that there is a significant interaction effect. Analyses were conducted for all six personality traits, for both the initial test scores and the retest scores, against all four in-basket performance dimensions.

A review of the initial results suggested that multicollinearity remained an issue; the tolerance values for many of the interaction terms were less than .10 (Field, 2013). To address this issue, each interaction term was squared with itself, and the analyses were conducted again. In addition, 9 cases were omitted from analyses due to their residual values (i.e., the residuals were more than three standard deviations from the mean; Field, 2013). In the subsequent analyses, tolerance values were above .10 and VIF values were less than 10.00 (Field, 2013).

In general, the results of the hypotheses were mixed. Regarding coordination, situational strength did not moderate any of the initial personality test–performance relationships. However, situational strength moderated the

relationship between personality and coordination, achievement and coordination and cautiousness and coordination. These results are summarized in Tables 6–7. Regarding decisiveness, the results were the same for both the initial personality test and the personality retest: situational strength moderated the relationship between competence and decisiveness, achievement striving and decisiveness, and between self-discipline and decisiveness. These results are summarized in Tables 8–9.

Regarding information management, the achievement–information management and self-discipline–information management relationships were moderated by situational strength for both the initial personality test and the retest. The competence–information management relationship was moderated by situational strength when retest scores were used. These results are summarized in Tables 10–11. Finally, regarding problem-awareness, the achievement–striving–problem awareness and self-discipline–problem awareness relationships were moderated by situational strength for both the initial personality test and the retest. The duty–problem awareness relationship was moderated when the initial test scores were used, and the cautiousness–problem awareness relationship was moderated when the retest scores were used. These results are summarized in Tables 12–13.

Table 6

Situational Strength as a Moderator of the Initial Personality Test–Coordination Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.010	0.010**
Competence	0.26	0.12	0.07		
Situational Strength	-0.25	0.11	-0.08		
Step 2				0.013	0.003
Competence	0.42	0.15	0.11		
Situational Strength	-0.23	0.11	-0.07		
Moderator	-0.06	0.03	-0.07		
Order					
Step 1				0.005	0.005
Order	0.03	0.09	0.01		
Situational Strength	-0.25	0.12	-0.08		
Step 2				0.006	0.001
Order	0.09	0.11	0.04		
Situational Strength	-0.23	0.12	0.07		
Moderator	-0.03	0.03	-0.04		

Situational Strength as a Moderator of the Initial Personality Test–Coordination Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.009	0.009*
Duty	0.26	0.13	0.07		
Situational Strength	-0.26	0.11	-0.08		
Step 2				0.009	0.000
Duty	0.25	0.18	0.06		
Situational Strength	-0.26	0.11	-0.08		
Moderator	0.00	0.05	0.00		
Achievement					
Step 1				0.011	0.011**
Achievement	0.27	0.11	0.08		
Situational Strength	-0.24	0.11	-0.07		
Step 2				0.013	0.003
Achievement	-0.44	0.15	0.13		
Situational Strength	-0.21	0.11	-0.07		
Moderator	-0.04	0.03	-0.08		

Situational Strength as a Moderator of the Initial Personality Test–Coordination Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.012	0.012**
Discipline	0.27	0.10	0.09		
Situational Strength	-0.25	0.11	-0.08		
Step 2				0.013	0.001
Discipline	0.28	0.14	0.12		
Situational Strength	-0.23	0.11	-0.07		
Moderator	-0.03	0.02	-0.05		
Cautiousness					
Step 1				0.010	0.010**
Cautiousness	0.25	0.11	0.08		
Situational Strength	-0.26	0.11	-0.08		
Step 2				0.011	0.001
Cautiousness	0.18	0.14	0.06		
Situational Strength	-0.27	0.11	-0.08		
Moderator	0.03	0.04	0.03		

Note. $N = 921$–938. Situational Strength = dummy-coded moderator variable. Moderator = personality facet x moderator interaction term. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are due to rounding error.

Table 7

Situational Strength as a Moderator of the Personality Retest–Coordination
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.007	0.007*
Competence	0.17	0.12	0.05		
Situational Strength	-0.24	0.11	-0.07		
Step 2				0.011	0.004*
Competence	0.37	0.16	0.10		
Situational Strength	-0.20	0.11	-0.06		
Moderator	-0.07	0.03	-0.09		
Order					
Step 1				0.008	0.008*
Order	0.15	0.09	0.06		
Situational Strength	-0.28	0.11	-0.08		
Step 2				0.008	0.000
Order	0.17	0.11	0.07		
Situational Strength	-0.27	-0.12	-0.08		
Moderator	-0.01	0.03	-0.01		

Situational Strength as a Moderator of the Personality Retest–Coordination
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.011	0.011**
Duty	0.33	.013	0.08		
Situational Strength	-0.26	0.11	-0.08		
Step 2				0.014	0.003
Duty	0.54	0.18	0.13		
Situational Strength	-0.23	0.11	0.13		
Moderator	-0.09	0.05	-0.08		
Achievement					
Step 1				0.009	0.009**
Achievement	0.23	0.11	0.07		
Situational Strength	-0.24	0.11	-0.07		
Step 2				0.015	0.006*
Achievement	0.46	0.15	0.14		
Situational Strength	-0.20	0.11	-0.06		
Moderator	-0.06	0.03	-0.10		

Situational Strength as a Moderator of the Personality Retest–Coordination Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline				0.008	0.008*
Step 1					
Discipline	0.20	0.10	0.06		
Situational Strength	-0.23	0.11	-0.07		
Step 2				0.011	0.003
Discipline	0.36	0.14	0.11		
Situational Strength	-0.20	0.11	-0.06		
Moderator	-0.40	0.03	-0.07		
Cautiousness					
Step 1				0.006	0.006
Cautiousness	0.12	0.11	0.04		
Situational Strength	-0.23	0.11	-0.07		
Step 2				0.014	0.008**
Cautiousness	0.43	0.15	0.13		
Situational Strength	-0.19	0.11	-0.06		
Moderator	-0.01	0.04	-0.13		

Note. $N = 921–938$. Situational Strength = dummy-coded moderator variable. Moderator = personality facet x moderator interaction term. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are due to rounding error.

Table 8

Situational Strength as a Moderator of the Initial Personality Test–Decisiveness
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.051	0.051**
Competence	1.06	0.15	0.23		
Situational Strength	-0.21	-0.14	-0.05		
Step 2				0.056	0.005*
Competence	1.31	0.19	0.28		
Situational Strength	-0.18	0.14	-0.04		
Moderator	-0.08	0.04	-0.09		
Order					
Step 1				0.016	0.016**
Order	0.44	0.11	0.14		
Situational Strength	-0.35	0.15	-0.08		
Step 2				0.018	0.002
Order	0.53	0.14	0.17		
Situational Strength	-0.32	-0.15	-0.08		
Moderator	-0.04	0.04	-0.05		

Situational Strength as a Moderator of the Initial Personality Test–Decisiveness
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.046	0.046**
Duty	1.09	0.16	0.22		
Situational Strength	-0.22	0.14	-0.05		
Step 2				0.046	0.000
Duty	0.99	0.23	0.20		
Situational Strength	-0.24	0.14	-0.06		
Moderator	0.04	0.06	0.03		
Achievement					
Step 1				0.046	0.046**
Achievement	0.92	0.14	0.22		
Situational Strength	-0.18	0.14	-0.04		
Step 2				0.052	0.006*
Achievement	1.24	0.19	0.29		
Situational Strength	-0.13	-0.14	-0.03		
Moderator	-0.08	0.04	-0.11		

Situational Strength as a Moderator of the Initial Personality Test–Decisiveness
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.045	0.045**
Discipline	0.84	0.13	0.21		
Situational Strength	-0.19	0.14	-0.05		
Step 2				0.053	0.008**
Discipline	1.19	0.18	0.30		
Situational Strength	-0.14	0.14	-0.03		
Moderator	-0.09	0.03	-0.13		
Cautiousness					
Step 1				0.050	0.050**
Cautiousness	0.94	0.13	0.23		
Situational Strength	-0.26	0.14	-0.06		
Step 2				0.050	0.000
Cautiousness	0.95	0.18	0.23		
Situational Strength	-0.26	0.14	-0.06		
Moderator	-0.01	0.05	-0.01		

Note. $N = 921–938$. Situational Strength = dummy-coded moderator variable.
Moderator = personality facet x moderator interaction term. B = unstandardized
coefficient; SE B = standard error of the unstandardized coefficient; ß =
standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are
due to rounding error.

Table 9

Situational Strength as a Moderator of the Personality Retest–Decisiveness
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.050	0.050**
Competence	1.02	0.15	0.22		
Situational Strength	-0.17	0.14	-0.04		
Step 2				0.056	0.006*
Competence	1.21	0.19	0.29		
Situational Strength	-0.12	-0.14	-0.03		
Moderator	-0.10	0.04	-0.10		
Order					
Step 1				0.030	0.030**
Order	0.59	0.11	0.18		
Situational Strength	-0.33	0.14	-0.08		
Step 2				0.030	0.000
Order	0.67	0.14	0.20		
Situational Strength	-0.31	0.15	-0.07		
Moderator	-0.03	0.04	-0.04		

Situational Strength as a Moderator of the Personality Retest–Decisiveness
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.054	0.054**
Duty	1.19	0.16	0.23		
Situational Strength	-0.20	0.13	-0.05		
Step 2				0.054	0.000
Duty	1.13	0.23	0.26		
Situational Strength	-0.18	0.14	-0.04		
Moderator	-0.05	0.06	-0.04		
Achievement					
Step 1				0.042	0.042**
Achievement	0.87	0.14	0.20		
Situational Strength	-0.16	-0.14	-0.04		
Step 2				0.048	0.006**
Achievement	1.18	0.19	0.28		
Situational Strength	-0.11	0.14	-0.03		
Moderator	-0.08	0.03	-0.11		

Situational Strength as a Moderator of the Personality Retest–Decisiveness Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.051	0.051**
Discipline	0.92	0.13	0.23		
Situational Strength	-0.13	0.13	-0.03		
Step 2				0.065	0.015**
Discipline	1.42	0.18	0.35		
Situational Strength	-0.05	0,14	-0.01		
Moderator	-0.12	0.03	-0.17		
Cautiousness					
Step 1				0.052	0.052**
Cautiousness	0.96	0.13	0.23		
Situational Strength	-0.20	0.14	-0.05		
Step 2				0.054	0.002
Cautiousness	1.10	0.19	0.26		
Situational Strength	-0.18	0.14	-0.04		
Moderator	-0.05	0.04	-0.05		

Note. N = 921–938. Situational Strength = dummy-coded moderator variable. Moderator = personality facet x moderator interaction term. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are due to rounding error.

Table 10

Situational Strength as a Moderator of the Initial Personality Test–Information Management Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.073	0.073**
Competence	1.41	0.17	0.27		
Situational Strength	-0.16	0.15	-0.04		
Step 2				0.076	0.003
Competence	1.62	0.21	0.31		
Situational Strength	-0.13	0.15	-0.03		
Moderator	-0.07	0.04	-0.07		
Order					
Step 1				0.022	0.022**
Order	0.57	0.13	0.16		
Situational Strength	-0.34	0.17	-0.07		
Step 2				0.025	0.003
Order	0.73	0.15	0.21		
Situational Strength	-0.30	0.17	-0.06		
Moderator	-0.08	0.04	-0.08		

Situational Strength as a Moderator of the Initial Personality Test–Information Management Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.095	0.095**
Duty	1.76	0.18	0.31		
Situational Strength	-0.24	0.15	-0.05		
Step 2				0.095	0.000
Duty	1.75	0.24	0.31		
Situational Strength	-0.24	0.15	-0.05		
Moderator	0.01	0.06	0.00		
Achievement					
Step 1				0.072	0.072**
Achievement	1/28	0.15	0.27		
Situational Strength	-0.14	0.15	-0.03		
Step 2				0.078	0.006**
Achievement	1.65	0.21	0.35		
Situational Strength	-0.08	0.15	-0.02		
Moderator	-0.10	0.04	-0.11		

Situational Strength as a Moderator of the Initial Personality Test–Information
Management Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.065	0.065**
Discipline	1.13	0.14	0.26		
Situational Strength	-0.14	0.15	-0.03		
Step 2				0.076	0.011**
Discipline	1.58	0.19	0.36		
Situational Strength	-0.07	0.15	-0.02		
Moderator	-0.11	0.03	-0.15		
Cautiousness					
Step 1				0.083	0.083**
Cautiousness	1.35	0.15	0.29		
Situational Strength	-0.26	0.15	-0.06		
Step 2				0.084	0.001
Cautiousness	1.49	0.20	0.32		
Situational Strength	-0.24	0.15	-0.05		
Moderator	-0.06	0.05	-0.05		

Note. $N = 921–938$. Situational Strength = dummy-coded moderator variable.
Moderator = personality facet x moderator interaction term. B = unstandardized
coefficient; SE B = standard error of the unstandardized coefficient; ß =
standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are
due to rounding error.

Table 11

Situational Strength as a Moderator of the Personality Retest–Information Management Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.080	0.080**
Competence	1.43	0.16	0.28		
Situational Strength	-0.15	0.15	-0.03		
Step 2				0.085	0.005*
Competence	1.76	0.21	0.35		
Situational Strength	-0.09	0.15	-0.02		
Moderator	-0.11	0.05	-0.10		
Order					
Step 1				0.029	0.029**
Order	0.66	0.13	0.18		
Situational Strength	-0.27	0.16	-0.06		
Step 2				-.029	0.000
Order	0.72	0.16	0.20		
Situational Strength	-0.25	0.16	-0.05		
Moderator	-0.03	0.04	-0.03		

Situational Strength as a Moderator of the Personality Retest–Information Management Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.103	0.103**
Duty	1.84	0.18	0.32		
Situational Strength	-0.19	0.15	-0.04		
Step 2				0.104	0.001
Duty	2.03	0.25	0.36		
Situational Strength	-0.17	0.15	-0.04		
Moderator	-0.07	0.06	-0.05		
Achievement					
Step 1				0.072	0.072**
Achievement	1.28	0.15	0.27		
Situational Strength	-0.12	0.15	-0.03		
Step 2				0.079	0.007**
Achievement	1.63	0.20	0.34		
Situational Strength	-0.06	0.15	-0.01		
Moderator	-0.10	0.04	-0.11		

Situational Strength as a Moderator of the Personality Retest–Information Management Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.073	0.073**
Discipline	1.24	0.14	0.27		
Situational Strength	-0.08	0.15	-0.02		
Step 2				0.089	0.016**
Discipline	1.81	0.20	0.40		
Situational Strength	0.02	0.15	0.00		
Moderator	-0.14	0.04	-0.18		
Cautiousness					
Step 1				0.101	0.101**
Cautiousness	1.49	0.15	0.32		
Situational Strength	-0.17	0.15	-0.04		
Step 2				0.101	0.000
Cautiousness	1.46	0.21	0.32		
Situational Strength	-0.18	0.15	-0.04		
Moderator	0.01	0.05	0.01		

Note. $N = 921–938$. Situational Strength = dummy-coded moderator variable. Moderator = personality facet x moderator interaction term. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are due to rounding error.

Table 12

Situational Strength as a Moderator of the Initial Personality Test–Problem Awareness Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.100	0.100**
Competence	1.61	0.16	0.32		
Situational Strength	-0.02	0.14	0.00		
Step 2				0.103	0.003
Competence	1.79	0.20	0.35		
Situational Strength	0.01	0.15	0.00		
Moderator	-0.06	0.04	-0.06		
Order					
Step 1				0.014	0.014**
Order	0.44	0.13	0.13		
Situational Strength	-0.11	0.17	-0.02		
Step 2				0.015	0.001
Order	0.41	0.15	0.12		
Situational Strength	-0.12	0.17	-0.03		
Moderator	0.01	0.04	0.02		

Situational Strength as a Moderator of the Initial Personality Test–Problem Awareness Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.103	0.103&&
Duty	1.77	0.17	0.32		
Situational Strength	-0.05	0.14	-0.01		
Step 2				0.109	0.006**
Duty	1.38	0.23	0.25		
Situational Strength	-0.11	0.15	-0.02		
Moderator	0.17	0.07	0.11		
Achievement					
Step 1				0.102	0.012**
Achievement	1.50	0.15	0.32		
Situational Strength	0.02	0.14	0.00		
Step 2				0.106	0.004*
Achievement	1.77	0.20	0.38		
Situational Strength	0.06	0.15	0.01		
Moderator	-0.07	0.04	-0.09		

Situational Strength as a Moderator of the Initial Personality Test–Problem Awareness Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.099	0.099**
Discipline	1.36	0.14	0.31		
Situational Strength	0.01	0.14	0.00		
Step 2				0.106	0.007**
Discipline	1.72	0.19	0.40		
Situational Strength	0.06	0.15	0.01		
Moderator	-0.09	0.03	-0.12		
Cautiousness					
Step 1				0.092	0,092**
Cautiousness	1.39	0.14	0.31		
Situational Strength	-0.08	0.15	-0.02		
Step 2				0.096	0.004
Cautiousness	1.15	0.19	0.26		
Situational Strength	-0.11	0.15	-0.02		
Moderator	0.09	0.05	0.08		

Note. $N = 921–938$. Situational Strength = dummy-coded moderator variable. Moderator = personality facet x moderator interaction term. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are due to rounding error.

Table 13

Situational Strength as a Moderator of the Personality Retest–Problem Awareness
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.106	0.106**
Competence	1.63	0.16	0.32		
Situational Strength	0.04	0.14	0.01		
Step 2				0.106	0.000
Competence	1.75	0.21	0.35		
Situational Strength	0.06	0.15	0.01		
Moderator	-0.04	0.05	-0.05		
Order					
Step 1				0.038	0.038**
Order	0.73	0.12	0.20		
Situational Strength	-0.13	0.15	-0.03		
Step 2				0.038	0.000
Order	0.66	0.16	0.19		
Situational Strength	-0.14	0.16	-0.03		
Moderator	0.02	0.04	0.03		

Situational Strength as a Moderator of the Personality Retest–Problem Awareness
Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.104	0.104**
Duty	1.84	0.18	0.32		
Situational Strength	-0.03	0.14	-0.01		
Step 2				0.107	0.003
Duty	1.54	0.25	0.27		
Situational Strength	-0.08	0.15	-0.02		
Moderator	0.13	0.08	0.08		
Achievement				0.096	0.096**
Step 1					
Achievement	1.45	0.15	0.31		
Situational Strength	0.04	0.14	0.01		
Step 2				0.099	0.003*
Achievement	1.71	0.20	0.36		
Situational Strength	0.08	0.15	0.02		
Moderator	-0.07	0.04	-0.08		

Situational Strength as a Moderator of the Personality Retest–Problem Awareness Relationship

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.092	0.092**
Discipline	1.37	0.14	0.30		
Situational Strength	0.09	0.14	0.02		
Step 2				0.101	0.009**
Discipline	1.78	0.20	0.39		
Situational Strength	0.16	0.15	0.03		
Moderator	-0.11	0.04	-0.13		
Cautiousness					
Step 1				0.102	0.102**
Cautiousness	1.46	0.14	0.32		
Situational Strength	-0.01	0.14	0.00		
Step 2				0.109	0.007**
Cautiousness	1.06	0.21	0.23		
Situational Strength	-0.06	0.15	-0.01		
Moderator	0.13	0.05	0,12		

Note. $N = 921–938$. Situational Strength = dummy-coded moderator variable. Moderator = personality facet x moderator interaction term. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$. Any discrepancies in ΔR^2 are due to rounding error.

Overall, the results suggest that under certain conditions, situational strength moderates the relationship between facet-level personality and performance on a job-relevant criterion. It should be noted that these results should be interpreted cautiously. While efforts were made to address issues regarding multicollinearity and extreme cases, a visual analysis of the P-P plots and residual plots suggest that for some of the analyses, there was a minor degree of heteroskedasticity.

Hypotheses XIII–XVIII

The third set of hypotheses state that the initial test correlations between facet-level conscientiousness and criterion performance will not be significantly different from the retest correlations between facet-level conscientiousness and criterion performance. As with the first set of hypotheses, these hypotheses were assessed by using Steiger's (1980) test for dependent correlations (i.e., the criterion performance score is the same for both correlations). The analyses were conducted by using Hoerger's (no date) on-line calculator. The four data points that were used were the correlations between the initial facet-level scores and the criterion performance dimensions, the correlations between retest facet-level scores and the criterion performance dimensions, the correlations between the initial and retest facet-level scores, and the sample sizes. Separate analyses were conducted for each performance dimension.

The results of the analyses fell into two categories (i.e., the hypotheses were either not supported or supported). The hypotheses were not supported for the order facet for the performance dimensions of coordination, decisiveness, and

problem awareness; likewise, the hypotheses were not supported for the

achievement striving facet for problem awareness. For the hypotheses that were

not supported, there was a statistically significant difference between the initial

test and retest correlations with the performance dimensions; in all of these cases,

the retest scores were higher than the initial scores. In all other instances, the

hypotheses were supported (i.e., the differences between the initial personality

test–performance correlations and the personality retest–performance correlations

were not significant), suggesting that personality retest effects do not have a

significant impact on validity. The results are summarized in Table 14.

Table 14

Differences Between Initial and Retest Facet–Criterion Correlations

Facet–Performance Dimension	Z Score*	Significance
Competence–Coordination	1.48	0.14
Competence–Decisiveness	-0.38	0.70
Competence–Information Management	-0.47	0.64
Competence–Problem Awareness	-0.09	0.93
Order–Coordination	-1.93	**0.05**
Order–Decisiveness	-2.32	**0.02**
Order–Information Management	-1.00	0.32
Order–Problem Awareness	-3.07	**0.02**
Duty–Coordination	-0.09	0.37
Duty–Decisiveness	-1.01	0.31
Duty–Information Management	-0.50	0.62
Duty–Problem Awareness	-0.54	0.59

Differences Between Pretest and Posttest Facet–Criterion Correlations

Facet–Performance Dimension	*Z* Score*	Significance
Achievement–Coordination	0.49	0.62
Achievement–Decisiveness	0.45	0.65
Achievement–Information Management	0.10	0.92
Achievement–Problem Awareness	-3.66	**<0.01**
Discipline–Coordination	0.60	0.49
Discipline–Decisiveness	-0.85	0.39
Discipline–Information Management	-1.07	0.29
Discipline–Problem Awareness	0.46	0.64
Caution–Coordination	1.41	0.16
Caution–Decisiveness	-0.44	0.66
Caution–Information Management	-1.40	0.16
Caution–Problem Awareness	-0.74	0.46

Note. N varies from 906 to 926 due to missing values. *Positive scores indicate that pretest facet-level personality scores were higher than posttest facet-level personality scores. Correlations that are significantly different from one another are set in bold type.

Research Questions I–VI

When examining the personality–performance relationship, researchers generally assume the function is linear. When testing this assumption, Robie and Ryan (1999) found did not find evidence supporting either a quadradic or a cubic effect. However, a few recent studies contradict this finding. At least two teams of researchers have found that the relationship between conscientiousness and performance is nonlinear (LaHuis et al., 2005, Le et al., 2011).

The question regarding linearity was examined by using hierarchical regression. In the first step, the in-basket performance dimension was regressed onto facet-level personality. In the second step, a quadradic product term was added as a predictor, and in the third step, a cubic term was added as a predictor (Cohen, Cohen, West, & Aiken, 2003). Facet-level personality scores were centered around the mean before the product terms were created to reduce collinearity effects (LaHuis et al., 2005). A statistically significant change in R^2 suggests that the relationship between the predictor and the criterion is nonlinear. Separate regression analyses were conducted for each personality facet and for each in-basket performance dimension. Given that this was a research question, the analyses were conducted using the initial scores (i.e., separate analyses were not conducted using the retest scores as well).

The results of the analyses were mixed. There were some instances in which none of the predictor–criterion regression models were significant, regardless of the form of the relationship (e.g., competence–coordination). There were some instances in which only the linear model was significant (e.g., order–decisiveness) or only the quadradic model was significant (e.g., duty–coordination). Likewise, there were some instances in which all three regression models were significant (e.g., achievement striving–problem solving). Taken together, the results, which are summarized in Tables 15–18 suggest that it is inappropriate to assume that the personality–performance relationship is always linear. Curve estimates of the significant nonlinear relationships can be found in Appendix F.

The quadradic relationships between duty and coordination and between cautiousness and coordination are subtle *U*s in which higher and lower facet scores are associated with a higher degree of coordination. Theoretically, these relationships do not make sense. The quadradic relationships between achievement and decisiveness and between discipline and decisiveness are subtle inverted *U*s in which moderate facet scores are associated with a higher degree of decisiveness. Theoretically, these relationships do not make sense. The quadradic relationship between duty and information management and between duty and problem awareness are subtle curved lines that looks like slightly curved positive regression lines; the model suggests that as duty scores increase, information management scores increase. These models align with our general assumption regarding predictor-criterion linearity. All the significant quadradic relationships are difficult to interpret. It also appears that they may be impacted by outliers. However, the cases that were excluded during the analyses of Hypotheses VII–XII were excluded from these analyses as well. A visual analysis of the nonsignificant nonlinear relationships provides clarity regarding their nonsignificance: The estimated quadradic and cubic curves deviate only slightly from the estimated linear regression line. In all cases, the slope of the linear regression was nearly horizontal.

Table 15

Form of the Relationship Between Facet-Level Predictors and Coordination

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.003	0.003
Linear Term	0.21	0.13	0.05		
Step 2				0.003	0.000
Linear Term	0.25	0.17	0.06		
Quadradic Term	-0.04	0.12	-0.16		
Step 3				0.008	0.005
Linear Term	0.08	0.19	0.02		
Quadradic Term	0.66	0.19	0.02		
Cubic Term	-0.23	0.10	-0.24		

Form of the Relationship Between Facet-Level Predictors and Coordination

Facet	B	SE B	ß	R^2	ΔR^2
Order					
Step 1				0.001	0.001
Linear Term	-0.09	0.09	-0.03		
Step 2				0.001	0.000
Linear Term	-0.12	0.10	-0.04		
Quadradic Term	0.05	0.11	0.02		
Step 3				0.001	0.000
Linear Term	-0.11	0.13	-0.04		
Quadradic Term	0.05	0.20	0.02		
Cubic Term	-0.01	0.13	-0.00		

Form of the Relationship Between Facet-Level Predictors and Coordination

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.001	0.001
Linear Term	0.16	0.14	0,04		
Step 2				0.005	0.004*
Linear Term	-0.15	0.21	-0.04		
Quadradic Term	0.37	0.19	0.10		
Step 3				0.006	0.000
Linear Term	-0.19	0.22	-0.04		
Quadradic Term	0.71	0.55	0.19		
Cubic Term	-0.18	0.27	-0.09		

Form of the Relationship Between Facet-Level Predictors and Coordination

Facet	B	SE B	ß	R^2	ΔR^2
Achievement					
Step 1				0.005	0.005*
Linear Term	0.25	0.11	0.07		
Step 2				0.006	0.001
Linear Term	0.36	0.20	0.10		
Quadradic Term	-0.09	0.12	-0.04		
Step 3				0.011	0.006*
Linear Term	0.08	0.23	0.02		
Quadradic Term	0.59	0.32	0.27		
Cubic Term	-0.21	0.09	-0.26		

Form of the Relationship Between Facet-Level Predictors and Coordination

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.005	0.005*
Linear Term	0.24	0.11	0.07		
Step 2				0.006	0.001
Linear Term	0.13	0.19	0.04		
Quadradic Term	0.08	0.10	0.04		
Step 3				0.014	0.008**
Linear Term	-0.18	0.22	-0.06		
Quadradic Term	0.89	0.31	0.48		
Cubic Term	-0.25	0.09	-0.38		

Form of the Relationship Between Facet-Level Predictors and Coordination

Facet	B	SE B	ß	R^2	ΔR^2
Cautiousness					
Step 1				0.003	0.003
Linear Term	0.19	0.11	0.05		
Step 2				0.012	0.009**
Linear Term	-0.21	0.18	-0.06		
Quadradic Term	0.37	0.13	0.15		
Step 3				0.012	0.000
Linear Term	-0.22	0.18	-0.07		
Quadradic Term	0.53	0.40	0.21		
Cubic Term	-0.07	0.17	-0.06		

Note. N varies from 932 to 948 due to missing values. Linear Model = personality facet; Quadradic Model = squared mean-centered personality facet; Cubic Model = cubed mean-centered personality facet. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$.

Table 16

Form of the Relationship Between Facet-Level Predictors and Decisiveness

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.047	0.047**
Linear Term	1.02	0.15	0.22		
Step 2				0.050	0.003
Linear Term	1.26	0.20	0.27		
Quadradic Term	-0.25	0.14	-0.08		
Step 3				0.057	0.007**
Linear Term	1.04	0.22	0.22		
Quadradic Term	0.69	0.39	0.21		
Cubic Term	-0.30	0.12	-0.27		

Form of the Relationship Between Facet-Level Predictors and Decisiveness

Facet	B	SE B	ß	R^2	ΔR^2
Order					
Step 1				0.010	0.010**
Linear Term	0.31	0.10	0.10		
Step 2				0.011	0.001
Linear Term	0.38	0.13	0.12		
Quadradic Term	-0.13	0.13	-0.04		
Step 3				0.011	0.000
Linear Term	0.34	0.16	0.11		
Quadradic Term	-0.21	0.24	-0.06		
Cubic Term	0.06	0.15	0.04		

Form of the Relationship Between Facet-Level Predictors and Decisiveness

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.042	0.042**
Linear Term	1.04	0.16	0.21		
Step 2				0.043	0.001
Linear Term	0.88	0.25	0.17		
Quadradic Term	0.20	0.22	0.04		
Step 3				0.043	0.000
Linear Term	0.88	0.26	0.17		
Quadradic Term	0.19	0.64	0.04		
Cubic Term	0.00	0.32	0.00		

Form of the Relationship Between Facet-Level Predictors and Decisiveness

Facet	B	SE B	ß	R^2	ΔR^2
Achievement					
Step 1				0.044	0.044**
Linear Term	0.90	0.14	0.21		
Step 2				0.052	0.008**
Linear Term	1.43	0.23	0.33		
Quadradic Term	-0.42	0.15	-0.15		
Step 3				0.056	0.004
Linear Term	1.16	0.27	0.27		
Quadradic Term	0.23	0.38	0.08		
Cubic Term	-0.21	0.11	-0.20		

Form of the Relationship Between Facet-Level Predictors and Decisiveness

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.042	0.042**
Linear Term	0.82	0.13	0.21		
Step 2				0.050	0.008**
Linear Term	1.32	0.22	0.33		
Quadradic Term	-0.34	0.12	0.33		
Step 3				0.052	0.002
Linear Term	1.15	0.25	0.29		
Quadradic Term	0.09	0.36	0.04		
Cubic Term	-0.13	0.11	-0.17		

Form of the Relationship Between Facet-Level Predictors and Decisiveness

Facet	B	SE B	ß	R^2	ΔR^2
Cautiousness					
Step 1				0.045	0.045**
Linear Term	0.88	0.13	0.21		
Step 2				0.046	0.001
Linear Term	0.79	0.21	0.19		
Quadradic Term	0.09	0.15	0.03		
Step 3				0.046	0.000
Linear Term	0.77	0.21	0.19		
Quadradic Term	0.25	0.47	0.08		
Cubic Term	-0.07	0.20	-0.05		

Note. N varies from 932 to 948 due to missing values. Linear Model = personality facet; Quadradic Model = squared mean-centered personality facet; Cubic Model = cubed mean-centered personality facet. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$.

Table 17

Form of the Relationship Between Facet-Level Predictors and Information
Management

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.067	0.067**
Linear Term	1.36	0.17	0.26		
Step 2				0.067	0.000
Linear Term	1.33	0.22	0.25		
Quadradic Term	0.03	0.16	0.01		
Step 3				0.093	0.026**
Linear Term	0.84	0.24	0.16		
Quadradic Term	2.10	0.42	0.57		
Cubic Term	-0.67	0.13	-0.53		

Form of the Relationship Between Facet-Level Predictors and Information
Management

Facet	B	SE B	ß	R^2	ΔR^2
Order					
Step 1				0.016	0.016**
Linear Term	0.45	0.12	0.13		
Step 2				0.017	0.001
Linear Term	0.52	0.14	0.15		
Quadradic Term	-0.13	0.15	-0.03		
Step 3				0.018	0.001
Linear Term	0.62	0.17	0.18		
Quadradic Term	0.09	0.27	0.02		
Cubic Term	-0.16	0.17	-0.08		

Form of the Relationship Between Facet-Level Predictors and Information Management

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.088	0.008**
Linear Term	1.70	0.18	0.30		
Step 2				0.092	0.004*
Linear Term	1.29	0.27	0.23		
Quadradic Term	0.48	0.25	0.09		
Step 3				0.093	0.002
Linear Term	1.21	0.28	0.21		
Quadradic Term	1.34	0.71	0.26		
Cubic Term	-0.45	0.35	-0.16		

Form of the Relationship Between Facet-Level Predictors and Information Management

Facet	B	SE B	ß	R^2	ΔR^2
Achievement					
Step 1				0.071	0.071**
Linear Term	1.29	0.15	0.27		
Step 2				0.075	0.004
Linear Term	1.69	0.26	0.35		
Quadradic Term	-0.32	0.17	-0.10		
Step 3				0.078	0.003
Linear Term	1.43	0.30	0.30		
Quadradic Term	0.31	0.42	0.10		
Cubic Term	-0.20	0.12	-0.17		

Form of the Relationship Between Facet-Level Predictors and Information Management

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.061	0.061**
Linear Term	1.12	0.14	0.25		
Step 2				0.063	0.002
Linear Term	1.35	0.24	0.30		
Quadradic Term	-0.16	0.14	-0.07		
Step 3				0.070	0.008**
Linear Term	0.95	0.28	0.21		
Quadradic Term	0.89	0.40	0.36		
Cubic Term	-0.32	0.12	-0.36		

Form of the Relationship Between Facet-Level Predictors and Information Management

Facet	B	SE B	ß	R^2	ΔR^2
Cautiousness					
Step 1				0.078	0.078**
Linear Term	1.31	0.15	0.28		
Step 2				0.080	0.002
Linear Term	1.09	0.23	0.23		
Quadradic Term	0.21	0.17	0.06		
Step 3				0.083	0.003
Linear Term	1.00	0.23	0.21		
Quadradic Term	1.06	0.51	0.31		
Cubic Term	-0.38	0.22	-0.25		

Note. N varies from 932 to 948 due to missing values. Linear Model = personality facet; Quadradic Model = squared mean-centered personality facet; Cubic Model = cubed mean-centered personality facet. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$.

Table 18

Form of the Relationship Between Facet-Level Predictors and Problem Awareness

Facet	B	SE B	ß	R^2	ΔR^2
Competence					
Step 1				0.098	0.098**
Linear Term	1.59	0.16	0.31		
Step 2				0.098	0.000
Linear Term	1.71	0.22	0.33		
Quadradic Term	-.011	0.15	-0.03		
Step 3				0.122	0.023**
Linear Term	1.25	0.23	0.25		
Quadradic Term	1.78	0.41	0.50		
Cubic Term	-0.61	0.12	-0.50		

Form of the Relationship Between Facet-Level Predictors and Problem Awareness

Facet	B	SE B	ß	R^2	ΔR^2
Order					
Step 1				0.014	0.014**
Linear Term	0.41	0.11	0.12		
Step 2				0.015	0.001
Linear Term	0.33	0.14	0.90		
Quadradic Term	0.15	0.15	0.09		
Step 3				0.016	0.001
Linear Term	0.33	0.17	0.10		
Quadradic Term	0.16	0.26	0.04		
Cubic Term	0.00	0.16	0.00		

Form of the Relationship Between Facet-Level Predictors and Problem Awareness

Facet	B	SE B	ß	R^2	ΔR^2
Duty					
Step 1				0.100	0.100**
Linear Term	1.76	0.17	0.32		
Step 2				0.111	0.011**
Linear Term	1.08	0.26	0.20		
Quadradic Term	0.80	0,24	0.16		
Step 3				0.114	0.003
Linear Term	0.98	0.27	0.18		
Quadradic Term	1.77	0.68	0.36		
Cubic Term	-0.51	0.33	-0.19		

Form of the Relationship Between Facet-Level Predictors and Problem
Awareness

Facet	B	SE B	ß	R^2	ΔR^2
Achievement					
Step 1				0.102	0.102**
Linear Term	1.51	0.15	0.32		
Step 2				0.106	0.004*
Linear Term	1.92	0.25	0.41		
Quadradic Term	-0.33	0.16	-0.11		
Step 3				0.114	0.008**
Linear Term	1.47	0.29	0.31		
Quadradic Term	0.75	0.40	0.25		
Cubic Term	-0.34	0.12	-030		

Form of the Relationship Between Facet-Level Predictors and Problem Awareness

Facet	B	SE B	ß	R^2	ΔR^2
Discipline					
Step 1				0.097	0.097**
Linear Term	1.36	0.14	0.31		
Step 2				0.101	0.004*
Linear Term	1.75	0.23	0.40		
Quadradic Term	-0.26	0.13	-0.11		
Step 3				0.111	0.010**
Linear Term	1.32	0.27	0.30		
Quadradic Term	0.86	0.38	0.36		
Cubic Term	-0.35	0.11	-0.40		

Form of the Relationship Between Facet-Level Predictors and Problem Awareness

Facet	B	SE B	ß	R^2	ΔR^2
Cautiousness					
Step 1				0.090	0.090**
Linear Term	1.36	0.14	0.30		
Step 2				0.097	0.007*
Linear Term	0.93	0.22	0.20		
Quadradic Term	0.41	0.16	0.13		
Step 3				0.104	0.007**
Linear Term	0.79	0.23	0.18		
Quadradic Term	1.73	0.50	0.53		
Cubic Term	-0.58	0.21	-0.40		

Note. N varies from 932 to 948 due to missing values. Linear Model = personality facet; Quadradic Model = squared mean-centered personality facet; Cubic Model = cubed mean-centered personality facet. B = unstandardized coefficient; SE B = standard error of the unstandardized coefficient; ß = standardized coefficient; * $p < 0.05$, ** $p < 0.01$.

Discussion

There were multiple goals to this study. One was to examine the relationship between personality, measured at the facet level, and performance. Another goal was to better understand this relationship by examining whether situational strength acted as a moderator. A third goal was to examine evidence for practice effects. Finally, a fourth goal was to assess whether the assumption regarding linearity holds true. Each of these goals in relation to the results of the study will be discussed in more depth below.

Facet- Versus Trait-Level Conscientiousness

Overwhelmingly, there was a lack of support for the research hypotheses regarding the predictive ability of facets versus traits. Specifically, there was only one instance in which facet-level personality, as opposed to trait-level personality, was more strongly correlated with the criterion: self-discipline was more strongly correlated with coordination than was conscientiousness. However, this was only true when the initial personality test scores were used. The remaining statistically significant results suggest that trait-level conscientiousness might be a better predictor of performance across multiple criteria.

There are a couple possible explanations for these findings. First, while conscientiousness has been found to predict performance across a number of jobs (Barrick & Mount, 1991), very few studies have used in-basket exercises as a criterion. Thus, it may be the case that personality, whether measured at the facet level or the trait level, is an especially effective predictor of in-basket performance. Related to this is the fact that a formal job analyses was not used to

link the predictor with the criterion (i.e., the predictor and the criterion were chosen based on previous research that had examined the general relationship between personality and job performance). Thus, this study underscores the importance of conducting a job analysis and using specific findings (i.e., do not generalize too much when using validity generalization) when linking predictors and criteria.

Given the relatively robust findings regarding conscientiousness, the outcome of this study is less likely a result of the insufficient predictive power of the construct. It may be more likely that the findings are an artifact of research design. Given that this was a lab study, participants may not have been sufficiently motivated to exhibit the level of maximum performance one might expect in a high-stakes selection context. There is anecdotal evidence for this, namely the number of participants who were excluded from the study due to inaccurate responses on the manipulation and attention checks. Thus, it is important to continue to examine predictor–criterion relationships using data collected from the field. Doing so will increase ecological validity and enable us to better generalize findings to actual selection contexts.

As was noted in the literature review, findings regarding the use of facets versus traits have been mixed. Warr et al. (2005) found evidence suggesting that facets are better predictors while Salsgo et al. (2013) found evidence suggesting that broad traits are better predictors. Unfortunately, this study adds to the conflicting body of knowledge. Thus, from a theoretical perspective, this study further underscores the importance of researching facet– versus trait–performance

relationships. Specific suggestions for future research include additional moderator identification and analyses to determine the specific conditions under which facets versus traits better predict performance. Another suggestion is to further examine multiple dimensions of performance. Operationalizing job performance at a more granular level will better enable researchers to theoretically and statistically link facets of personality with dimensions of performance.

Situational Strength as a Moderator

The findings regarding situational strength were mixed. Of the 48 analyses that were conducted (i.e., 6 facets, 4 performance dimensions, initial and retest scores), slightly fewer than half were statistically significant, suggesting that future research will help us better understand the conditions under which situational strength has an impact on the personality–performance relationship. It is worth noting that participants' facet scores clustered toward the lower end of the scale, suggesting the presence of floor effects. Thus, it is possible that effects of the situational strength manipulation may have been artificially constrained.

In this study, situational strength was most important when achievement striving and self-discipline were used as the predictors. There is no theoretical reason to believe that some of the predictor–criterion relationships should be moderated while others are not. It is worth pointing out that three facet–decisiveness relationships (i.e., competence, achievement striving, self-discipline) were moderated by situational strength. Further research may help explain why facet–decisiveness relationships are more likely to be moderated by situational

strength than other facet–performance outcome relationships. As was noted above, the findings should be interpreted with caution; visual analyses of the P-P plots and the residual plots suggest that, for some of the relationships that were analyzed, there were minor violations regarding the assumption of homoskedasticity.

There are likely a number of explanations for these findings. It may be the case, despite the suggestions of Judge and Zapata (2016), that situational strength does not help us better understand the personality–performance relationship. However, given the number of significant findings in this study, it is more likely the case that the inclusion of a single moderator is insufficient (i.e., the complexity of human behavior cannot be modeled by two variables). Thus, future research is needed to better specify the personality–performance relationship. Additional studies could include additional constructs, as well as examine other types of relationships (e.g., mediated relationships, moderated-mediated relationships). In addition, future research should further examine the criterion space. As has been acknowledged, job performance is multidimensional, but many validation studies operationalize performance by using supervisor ratings. It may be the case that situational strength is more important for specific dimensions of task performance or for criteria that have received less research attention (e.g., teamwork versus taskwork in team settings, organizational citizenship behaviors, counterproductive workplace behaviors).

As was the case with the first set of hypotheses, the findings regarding these hypotheses are likely also the result of design artifacts. Participant

motivation was likely an issue. In addition, it may be the case that a limited amount of information about a fictional organization did not facilitate trait activation (Tett & Burnett, 2003). In other words, the experimental manipulation may not have been strong enough. In addition to examining more complex relationships, future research could also include other signals that might strengthen the impact of the manipulation. For example, using a more media-rich realistic job preview as part of a study could give applicants better insight into which personality constructs are most job relevant and the degree to which those constructs can be expressed on the job. From an organizational perspective, providing signals about situational strength would be beneficial, as doing so would better enable applicants to assess person–job fit.

Retesting Effects

In general, the findings regarding retesting effects suggest that participants' scores did not change much between the initial administration of the measure and the subsequent administration of the measure (i.e., there were changes in only 5 of the 24 relationships examined). Interestingly the scores only changed for order and achievement-striving.

While these findings may appear to conflict with previous findings regarding the presence of retesting effects, it is possible that the results are due to research design. Many of the previous studies that have examined this topic included a condition in which participants were explicitly told to fake their responses, thus resulting in a large effect. An issue with this particular study is that there was a short time lag between test administrations. Thus, the results may

be an indirect assessment of memory more than anything else. Participant

motivation may have been an issue as well. While the participants were told that

their test scores did not result in a job offer, the nature of the study may not have

motivated participants to change their responses to better qualify for a job.

At the same time, personality is thought to be a relatively stable construct.

Thus, it is not terribly surprising that there was limited evidence of practice

effects. In addition, the personality assessment was administered using best

practices that have been identified in the literature. Specifically, a warning

appealing to participants good nature was used (i.e., "Thank you for taking the

time to fully read each item and to answer to the best of your ability. Your efforts

will help us determine whom to hire."), and the facets being assessed were not

identified, thereby ensuring item transparency. As a result, participants may not

have been able to identify the constructs being assessed or the desired relationship

between personality and criterion performance.

Linearity of the Personality–Performance Relationship

One of the most beneficial results of this study are the findings regarding

the form of the relationship between personality and criterion performance.

Traditionally, validation studies have assumed that the relationship is linear.

However, a violation of this assumption helps explain the relatively small

conscientiousness–performance relationships that have been reported in the

literature (LeHuis et al., 2015). This study calls the traditional assumption into

question.

Specifically, there were significant quadradic or cubic relationships for four of the six facets for both coordination and problem awareness; there were significant relationships for three of the six facets for decisiveness and information management. From another perspective, four of the discipline– criteria relationships, three of the competence–criteria, duty–criteria, and achievement–criteria relationships, and two of the cautiousness–criteria relationships were nonlinear; none of the order–criteria relationships were nonlinear. It is necessary to point out that the mixed findings reported in this study may be the result of the design artifacts that have been mentioned (e.g., participant motivation). They may also be the result of additional, unspecified moderators.

The results suggest that researchers cannot make accurate personality– performance conclusions without also examining the form of the relationship between personality and performance. This is consistent with the existing literature. Robie and Ryan (1999) did not find evidence suggesting a nonlinear conscientiousness–performance relationship; however, they used supervisor ratings as the criterion. The two teams of authors how have found evidence for a nonlinear relationship used more specific criteria, including perceptual speed and accuracy (LaHuis et al., 2005) and task performance, organizational citizenship behaviors, and counterproductive workplace behaviors (Le et al., 2011).

Taken in conjunction with those findings, this study further underscores the both the theoretical and practical importance of expanding the criterion space. In other words, it is no longer sufficient to automatically assume a linear

predictor–criterion relationship. Likewise, it may be insufficient to continue to use a single, unidimensional criterion, namely supervisors' ratings of performance. Thus, there is a clear need for additional research, as has been pointed out previously (Schmitt, 2014); further study can help us better model the complexity between humans' characteristics and job performance.

Implications

Given the number of inconclusive findings (i.e., none of the hypotheses were fully supported or fully rejected), it is difficult to suggest clear implications for practitioners. That said, one implication is that practitioners should conduct high-quality job analyses before developing selection systems. Doing so will better enable them to link specific predictors with specific dimensions of job performance. In addition, this will help ensure that any attempts at validity generalization are not too general.

Another implication for practitioners is that they would greatly contribute to the literature by allowing researchers to conduct field studies within their organizations. This will help address some of the participant motivation issues identified above, particularly if predictive designs, rather than concurrent designs, are used when validating selection systems. In addition, the use of field data will increase ecological validity, thereby allowing researchers to better generalize their findings.

This study provides a number of implications for researchers. Most importantly, the results of this study suggest that future research is necessary. An example of this is the need for designs that better model the complex nature of the

predictor–performance relationship, as discussed above and as suggested by previous researchers (e.g., Judge & Zapata, 2016; Schmitt, 2014). To this end, it will be beneficial for researchers to identify additional moderators of the personality–performance relationship. Examples of additional moderators include employees' relationships with their supervisors (e.g., if employees are being recruited and selected internally rather than externally), the role of the job within the organization (e.g., whether the job is a managerial position or a line job), and the degree to which the job facilitates the various components of the job characteristics model (i.e., skill variety, task significance, autonomy, feedback; Hackman & Oldham, 1975).

In addition, researchers should continue to further refine the criterion space by developing additional measures and by taking a more nuanced approach to performance dimensionality. The low reliability of the in-basket used in this study provides further support for the need to develop psychometrically sound criterion measures.

Another opportunity is to better specify the form of the relationship between personality and performance. It is likely that the relationship is not linear; in addition, it is possible that the relationship is not stable (i.e., there are likely within-job differences as a result of moderated relationships, and there are likely between job differences such that the form of relationship is different for different jobs). Finally, the findings regarding the nonlinear relationships may be of interest to both researchers and practitioners who are responsible for composing teams (i.e., developing models that account for nonlinear relationships

will enable practitioners to select team members who have the optimal levels of complementary and/or supplementary personality traits).

Limitations

There are a number of limitations to this study that must be acknowledged. First, while the study included a constrained number of constructs (i.e., personality, in-basket performance, and situational strength), the number of personality facets—which were measured twice—and the number of performance dimensions that were included in the study resulted in a large number of analyses. Statistically, this is problematic in that it may have resulted in familywise error.

Another limitation of this study is that it was a lab study. While lab studies are beneficial because they allow researchers to control for extraneous variables, they limit the generalizability of the findings due to their diminished ecological validity. A specific concern regarding this study is that the experimental manipulation may not have been strong enough. While there is some evidence that situational strength moderated the personality–performance relationship, the signals provided in the study may not have been strong enough to result in trait activation or suppression. Another specific concern regarding this study is that the participants may have been more motivated to earn the $3.00 payment than to contribute to the scientific literature. There is some anecdotal evidence of this (e.g., the number of participants who were excluded from the study due to inaccurate responses to manipulation and attention checks, the number of participants who contacted me after they were excluded to ask that they be paid regardless of the quality of their work).

From an applied perspective, a major limitation of this study is that it does not adequately mirror a true selection context. It is unlikely that a contemporary organization would only use personality as a predictor of performance. While focusing on personality enabled me to better investigate the personality– performance relationship, this approach inhibits generalizability. It may have been more effective to conduct a more constrained study in which fewer facets were used and in which at least another predictor (e.g., cognitive ability) was included in the study.

Conclusion

While the findings of this study are inconclusive, they suggest that personality is a useful predictor of job performance. Further research will help us identify the ways in which and the conditions under which this construct can best be used in applied settings.

References

Adair, C. (2014). *Interventions for addressing faking on personality assessments for employee selection: A meta-analysis* (Unpublished doctoral dissertation). DePaul University, Chicago, IL.

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revivail of test bias research in preemployment testing. *Journal of Applied Psychology, 95,* 648–680.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist, 36,* 1086–1093.

Anseel, F., & Lievens, F. (2006). Certainty as a moderator of feedback reactions? A test of the strength of the self-verification motive. *Journal of Occupational and Organizational Psychology, 79,* 533–551.

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality, 29,* 432–442.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO Model of personality structure. *Personality and Social Psychology Review, 11,* 150–166.

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology, 77,* 836–874.

Avis, J. M., Kudisch, J. D., & Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology, 17,* 87–105.

Bakker, A. B., Demerouti, E., & ten Brummelhuis, L. (2012). Work engagement, performance, and active learning: The role of conscientiousness. *Journal of Vocational Behavior, 80,* 555–564.

Barger, P., Behrend, T. S., Sharek, D. J., & Sinar, E. E. (2011). I–O and the crowd: Frequently asked questions about using Mechanical Turk for research. *The Industrial–Organizational Psychologist, 49*(2) 11–17.

Barney, J. B., & Wright, P. M. (1998). On becoming a strategic partner: The role of human resources in gaining competitive advantage. *Human Resource Management, 37,* 31–46.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26.

Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78,* 111–118.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81,* 261–272.

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to
more important matters *Human Performance, 18,* 359–372.

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.

Barron, L. G., Randall, J. G., Trent, J. D., Johnson, J. F., & Villado, A. J. (2017).
Big Five traits: Predictors of retesting propensity and score improvement.
International Journal of Selection and Assessment, 25, 138–148.

Beaty, J. C., Jr., Cleveland, J. N., & Murphy, K. R. (2001). The relation between
personality and contextual performance in "strong" versus "weak"
situations. *Human Performance, 14,* 125–148.

Benet-Martinez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and
ethnic groups: Multitrait multimethod analysis of the Big Five in Spanish
and English. *Journal of Personality and Social Psychology, 75,* 729–750.

Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in
the criterion-related validity of cognitive ability tests: A qualitative and
quantitative review. *Journal of Applied Psychology, 96,* 881–906.

Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in
performance versus fairness resulting from two cut-score strategies.
Personnel Psychology, 62, 835–863.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M.
A. (2006). A meta-analytic investigation of job applicant faking in

personality measures. *International Journal of Selection and Assessment, 14,* 317–335.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management, 36, 1065–1105.*

Borkenau, P., & Ostenforf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences, 11,* 515–524.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10,* 99–109.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6,* 3–5.

Chatman, J. A. (1991). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly, 36,* 459–484.

Chiaburu, D. S., Oh, I.-S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The Five Factor Model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology, 96,* 1140–1166.

Christiansen, N. D., Goffin, R. D., Johnstonm N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47,* 847–860.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.

Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review, 13,* 62–72.

Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12,* 887–898.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* New York, NY: Harper and Row.

Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology, 42,* 25–36.

Demerouti, E. (2006). Job characteristics, flow, and performance: The moderating role of conscientiousness. *Journal of Occupational Health Psychology, 11,* 266–280.

Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research, 16,* 149–170.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91,* 40–57.

Dunlop, P. D., Morrison, D. L., & Cordery, J. L. (2011). Investigating retesting effects in a personnel selection context. *International Journal of Selection and Assessment, 19,* 217–221.

Ellingson, J. E., Heggestad, E. D., & Makarius, E. E. (2012). Personality retesting for managing intentional distortion. *Journal of Personality and Social Psychology, 102,* 1063–1076.

Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (1st edition, pp. 244–276). New York, NY: Guilford Press.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Los Angeles, CA: SAGE.

Fisher, D. M., Cunningham, S., Kerr, A. J., & Allscheid, S. P. (2017). Contextualized personality measures in employee selection: Extending frame-of-reference research with job applicant samples. *International Journal of Selection and Assessment, 25,* 18–35.

Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology, 61,* 579–616.

Gellatly, I. R., & Irving, P. G. (2001). Personality, autonomy, and contextual performance of managers. *Human Performance, 14,* 231–245.

Ghiselli, E. E., & Barthol, R. P. (1953). The validity of personality inventories in the selection of employees. *Journal of Applied Psychology, 37,* 18–20.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36,* 341–355.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18,* 135–164.

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology, 60,* 159–170.

Halpert, J. A., Gerjerts, K. G., Miller, L., Lukasik, M., & Fritts, T. (2008, April). Personality tests used for selection: Practice effects. In J. Hausknecht (Chair), *Recent research on retesting and its implications for selection.* Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Hattrup, K., & Jackson, S. E. (1996). Learning about individual differences by taking situations seriously. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 507–547). San Francisco, CA: Jossey-Bass Publishers.

Hattrup, K., O'Connell, M. S., & Wingate, P. H. (1998). Prediction of multidimensional criteria: Distinguishing task and contextual performance. *Human Performance, 11,* 305–319.

Hauser, D. J., Paolacci, G., & Chandler, L. (in press). Common concerns with MTurk as a participant pool: Evidence and solutions. In F. Kardes, P. Herr, & N. Schwarz (Eds.), *Handbook in research methods in consumer psychology.* New York, NY: Taylor & Francis.

Hausknecht, J. P. (2010). Candidate persistence and personality test practice effects: Implications for staffing system management. *Personnel Psychology, 63,* 299–324.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92,* 373–385.

Hoerger, M. (n.d.). An updated version of Steiger's Z and web-based calculator for testing the statistical significance of the difference between dependent correlations [On-line calculator]. Retrieved from http://www.psychmike.com/dependent_correlations.php

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employee selection. *Journal of Applied Psychology, 92,* 1270–1285.

Holladay, C. L., David, E., & Johnson, S. K. (2013). Retesteing personality in employee selection: Implications of the context, sample, and setting. *Psychological Reports: Measures and Statistics, 112,* 486–501.

Hough, L. M. (1992). The "Big Five" personality variables—construct confusion: Description versus prediction. *Human Performance, 5,* 139–155.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11,* 209–244.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect

of response distortion on those validities. *Journal of Applied Psychology, 75,* 581–595.

Hough, L. M., & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. Kepir Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 1). Thousand Oaks, CA: SAGE.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment, 9,* 152–194.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–98.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85,* 869–879.

Jackson, D. N., Wrobleski, V. R., & Ashton, M. C., (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13,* 371–388.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. (Ed.), *Handbook of personality: Theory and research.* New York, NY: Guilford Press.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the Five-Factor Model of

personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98,* 875–925.

Judge, T. A., & Zapata, C. P. (2016). The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Fic personality traits in predicting job performance. *Academy of Management Journal, 58,* 1149–1179.

Katigbak, M. S., Church, A. T., & Akamine, T. X. (1996). Cross-cultural generalizability of personality dimensions: Relating indigenous and imported dimensions in two cultures. *Journal of Personality and Social Psychology, 70,* 99–114.

Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Krammer, G., Sommer M., & Arendasy, M. E. (2017). The psychometric costs of applicants' faking: Examining measurement invariance and retest correlations across response conditions. *Journal of Personality Assessment, 99,* 510–523.

LaHuis, D. M., Martin, N. R., & Avis, J. M. (2005). Investigating nonlinear conscientiousness–job performance relations for clerical employees. *Human Performance, 18,* 199–212.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational Mechanical Turk, and other

convenience samples. *Industrial and Organizational Psychology, 8*(2), 142–164.

Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking on online personality measures for selection. *Journal of Applied Psychology, 96,* 202–210.

Le, H., Oh, I.-S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology, 96,* 113–133.

LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology, 53,* 563–593.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58,* 981–1007.

Lievens, F., Harris, M. H., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88,* 476–488.

Lievens, F., Reeve, C. L., & Heggestad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology, 92,* 1672–1682.

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: Longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology, 101,* 847–861.

Mattern, K. D., & Patterson, B. F. (2013). Test of slope and intercept bias in college admissions: A response to Aguinis, Culpepper, & Pierce (2010). *Journal of Applied Psychology, 98,* 134–137.

Matton, N., Vautier, S., & Raufaste, E. (2011). Test-specificity of the advantage of retaking cognitive ability tests. *International Journal of Selection and Assessment, 19,* 11–17.

McCrae, R. R., and Costa, P. T., Jr. (1987). Validation of the Five-Factor Model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90.

McCrae, R. R., and Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509–516.

McManus, M. A., & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. *Personnel Psychology, 52,* 137–148.

Meyer, R. D., Dalal, R. S., & Bonaccio, S. (2009). A meta-analytic investigation into the moderating effects of situational strength on the conscientiousness–performance relationship. *Journal of Organizational Behavior, 30,* 1077–1102.

Meyer, R. D., Dalal, R. S., & Hemida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management, 36,* 121–140.

Meyer, R. D., Dalal, R. S., José, I. J., Hermida, R., Chen, T. R., Vega, R. P., . . . Khare, V. P. (2014). Measuring job-related situational strength and assessing its interactive effects with personality on voluntary work behavior. *Journal of Management, 40,* 1010–1041.

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 333–352). Hillsdale, NJ: Lawrence Erlbaum Associates. .

Moon, H., Hollenbeck, J. R., Marinova, S., & Humphrey, S. E. (2008). Beneath the surface: Uncovering the relationship between extraversion and organizational citizenship behavior through a facet approach. *International Journal of Selection and Assessment, 16,* 143–154.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60,* 1029–1049.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60,* 683–729.

Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resources Management, 13,* 153–200.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1999). The joint relationship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management, 25,* 707–721.

Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance, 18,* 343–357.

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.

Nunnally, J. C., & Bernstein, I. A. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Nyberg, A. J., Moliterno, T. P., Hale, D., Jr., & Lepak, D. P. (2012). Resource-based perspective on unit-level human capital: A review and integration. *Journal of Management, 40,* 316–346.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17,* 609–626.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11,* 245–269.

Oppenheimer, D. M., Meybis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45,* 867–872.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology, 48,* 775–802.

Oswald, F. L., & Hough, L. M. (2011). Personality and its assessment in organizations: Theoretical and empirical developments. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 2, pp. 153–184). Washington, DC: APA.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making 5,* 411–419.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61,* 153–172.

Ployhart, R. E., & Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *Academy of Management Review, 36,* 127–150.

Pyburn, K. M. Jr., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61,* 143–151.

Randall, J. G., & Villado, A. J. (2017). Take two: Sources and deterrents of score change in employment retesting. *Human Resource Management Review, 27,* 536–553.

Raymond, M. R., Neustel, S., & Anderson, D. (2007). Retest effects on identical and parallel forms in certification and licensure testing. *Personnel Psychology, 60,* 367–396.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80,* 721–730.

Reeve, C. L., & Lam, H. (2007). The relation between practice effects, test-taker characteristics and degree of *g* saturation. *International Journal of Testing, 7,* 225–242.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126,* 3–25.

Robertson, I., Gibbons, P., Baron, H., MacIver, R., & Nyfield, G. (1999). Understanding management performance. *British Journal of Management, 10,* 5–12.

Robie, C., & Ryan, A M. (1999). Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. *International Journal of Selection and Assessment, 7,* 157–169.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and

educational settings: A meta-analysis. *Personnel Psychology, 54,* 297–330.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16,* 155–180.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist, 56,* 302–318.

Salgado, J. F. (1997). The Five Factor Model of personality and job performance in the European community. *Journal of Applied Psychology, 82,* 30–43.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76,* 323–346.

Salgado, J. F., Moscoso, S., & Berges, A. (2013). Conscientiousness, its facets, and the prediction of job performance ratings: Evidence against the narrow measures. *International Journal of Selection and Assessment, 21,* 74–84.

Schleicher, D. J., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology, 95,* 603–617.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: why there cannot be a debate. *Human Performance, 15,* 187–210.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection

    methods in personnel psychology: Practical and theoretical implications of

    85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work:

    Occupational attainment and job performance. *Journal of Personality and*

    Social psychology, 86, 162–173.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-

    reference effects on personality scale scores and criterion-related validity.

    Journal of Applied Psychology, 80, 607620.

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective

    performance at work. *Annual Review of Organizational Psychology and*

    Organizational Behavior, 1, 45–65.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and*

    Psychological Measurement, 23, 153–158.

Smithikrai, C. (2008). Moderating effect of situational strength on the relationship

    between personality traits and counterproductive work behaviour. *Asian*

    Journal of Social Psychology, 11, 253–263.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix.

    Psychological Bulletin, 87, 245–251.

Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing

    differential effects for conscientiousness and its subtraits. *Journal of*

    Applied Psychology, 84, 959–968.

te Nijenhuis, J., & van der Flier, H. (2000). Differential prediction of immigrant versus majority group training performance using cognitive ability and personality measures. *International Journal of Selection and Assessment, 8,* 54–60.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessment anxiety with an emphasis on self-report. In A, Tuma & J, Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88,* 500–517.

Tett, R. P., Guterman, H. A., Bleir, A., & Murphy, P. J. (2000). Development and content validation of a "hyperdimensional" taxonomy of managerial competence. *Human Performance, 13,* 205–251.

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.

Tett, R. P., Menard, D., Guterman, H., & Beauregard, R. S. (2001, April). *Challenges in computerizing work sample measures: Testing the equivalence of an in-basket exercise*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality–job performance relationship. *Journal of Organizational Behavior, 24,* 335–356.

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (ASD-TR-61-97). Lackland Air Force Base, Tx: Aeronautical Systems Divisions, Personnel Laboratory.

Uniform Guidelines on Employee Selection Procedures, 43 Fed. Reg. 38,296–38,309 (1978).

Van Iddekinge, C. H., Morgeson, F. P., Schleicher, D. J., & Campion, M. A. (2011). Can I retake it? Exploring subgroup differences and criterion-related validity in promotion retesting. *Journal of Applied Psychology, 96,* 941–955.

Villado, A. J., Randall, J. G., & Zimmer, C. U. (2016). The effects of method characteristics on retest score gains and criterion-related validity. *Journal of Business Psychology, 31,* 233–248.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S., III., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83,* 586–597.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59,* 197–210.

Walmsley, P. T., & Sackett, P. R. (2013). Factors affecting potential personality retest improvements after initial failure. *Human Performance, 26,* 390–408.

Warr, P., Bartram, D., & Martin, T. (2005). Personality and sales performance: Situational variation and interactions between traits. *International Journal of Selection and Assessment, 13,* 87–91.

Weiss, H. M., & Adler, S. (1984). Personality and organizational behavior. *Research in Organizational Behavior, 6,* 1–50.

Witt, L. A., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology, 87,* 164–169.

Wortman, J., Lucas, R. E., & Donnellan, M. B. (2012). Stability and change in the Big Five personality domains: Evidence from a longitudinal study of Australians. *Psychology and Aging, 27,* 867–874.

Appendix A

Personality Pretest and Retest

Thank you for applying for a job with Apex. As part of the hiring process, please complete the following personality scale. Thank you for taking the time to fully read each item and to answer to the best of your ability. Your efforts will help us determine whom to hire. [A 5-point scale (1 = *Extremely Accurately*, 5 = *Not accurately at all*) was provided and items marked with an asterisk were reverse scored.]

Competence

1. I complete work tasks successfully.

2. I excel in what I do at work.

3. I come up with good solutions to problems at work.

4. I know how to get things done at work.

5. I misjudge situations at work.*

6. I don't understand things at work.*

7. I have little to contribute at work.*

Orderliness

1. I like order at work.

2. I like to tidy up at work.

3. I want everything to be "just right" at work.

4. I love order and regularity at work.

5. I do things at work according to plan.

6. At work, I often forget to put things back in their proper place.*

7. I leave a mess in my workplace.*

Dutifulness

1. I try to follow the rules at work.

2. I keep my promises at work.

3. I tell the truth at work.

4. I listen to my conscience at work.

5. I break rules at work.*

6. I get others to do my work duties.*

7. At work, I do the opposite of what is asked.*

Achievement-striving

1. I work hard at work.

2. At work, I turn plans into action.

3. At work, I do more than what's expected of me.

4. At work, I set high standards for myself.

5. At work I am NOT highly motivated to succeed.*

6. At work, I do just enough to get by.*

7. I put little time and effort into my work.*

Self-discipline

1. I get work tasks done right away.

2. I carry out my plans at work.

3. I find it difficult to get down to work.*

4. I waste my time at work.*

5. At work, I need a push to get started.*

6. I have difficulty starting my work tasks. **

7. At work, I postpone decisions.*

Cautiousness

1. I avoid mistakes at work.

2. At work, I choose my words with care.

3. At work, I jump into things without thinking.*

4. I like to rush into things at work.*

5. I like to do crazy things at work.*

6. I like to act without thinking at work.*

7. At work, I often make last-minute plans.*

Appendix B

Criterion

Thank you for completing the personality measure.

Continue to imagine that you are applying for a job at Apex. As part of the hiring process, you have been asked to complete another task, which is described below. Imagine that you have recently been appointed plant manager of the Chicago branch of Duron Paints. Prior to this, you spent five years working as the operations manager at the Minneapolis branch of Duron Paints. You just began your first week of work at the Chicago branch. Here is some background information for you. Duron Paints produces paints and finishing products for houses and vehicles. The company is headquartered in Detroit, Michigan, but there are seven other branches throughout the country.

Your former boss

Ben Green was the plant manager in Minneapolis for many years. He was the person who introduced you to Duron Paints and was your mentor during your first years at the company.

Your new boss

David Burnam was the previous plant manager in the Chicago branch (i.e., your predecessor). He has been promoted to Vice-President of Manufacturing at Duron's Detroit headquarters. You used to work with David in Minneapolis and have always had a good relationship with him.

Your new colleagues

You just arrived in Chicago, and you haven't had time to meet your new

colleagues. Luckily, you have access to an organizational chart. Four managers report directly to you. The only person you already know is Gina Winters. She is your administrative assistant.

Recent events

Today is Sunday, December 16, and it is just before 3:00 PM. You were informed about a week and a half ago about your promotion. The day before yesterday, you left Minneapolis and arrived late the same day in Chicago.

David Burnam, your new boss, has scheduled a meeting for all the Duron Plant Managers (i.e., the managers from each branch across the country) for 10:00 AM tomorrow, Monday, in Detroit. To avoid the morning rush hour, your administrative assistant, Gina Winters, arranged a flight for you that is leaving at 6:30 this evening. You expect to be back in Chicago on Tuesday, December 18, at 2:00 PM.

Your task

At the moment, you have one hour before a ride share comes to pick you up for the airport. You want to use this time to catch up on emails. You have 10 messages. Each one contains a problem that concerns you. It is now your task to answer all of those e-mails before you leave for Detroit.

You'll have to work fairly quickly, and you should not spend any longer than 30 minutes on this task. Like every good manager, you need to try to answer every e-mail. Your most important task as plant manager is to coordinate the work of your employees so that the branch functions efficiently. Given that this is your last

chance to look at your e-mails before you get back from Detroit, you have to be very explicit if necessary.

Instructions

For each message, you are presented with four possible responses. Please use a 5-point scale to rate how likely it is that you would use each response option. In other words, please rate how likely you would be to use all four response options for all 10 emails.

You can answer the e-mails in the order you choose. The only requirement is that you give your opinion about each possible response.

Attention check

Before you complete this task, please answer the following items.

1. What is your role within the organization?

   Response options: administrative assistant, plant manager, vice president of manufacturing

2. What is the name of your new boss?

   Response options: Ben Green, David Burnam, Gina Winters

3. Your task is to rate how likely you would use each response option

   Response options: True, False

Organizational chart

Here is an organizational chart for Duron Paints. You may wish to right click on the image and open it in another tab so that you can reference it while completing the task. Remember that you report to David Burnam, the Vice-President of Manufacturing at the Detroit headquarters.
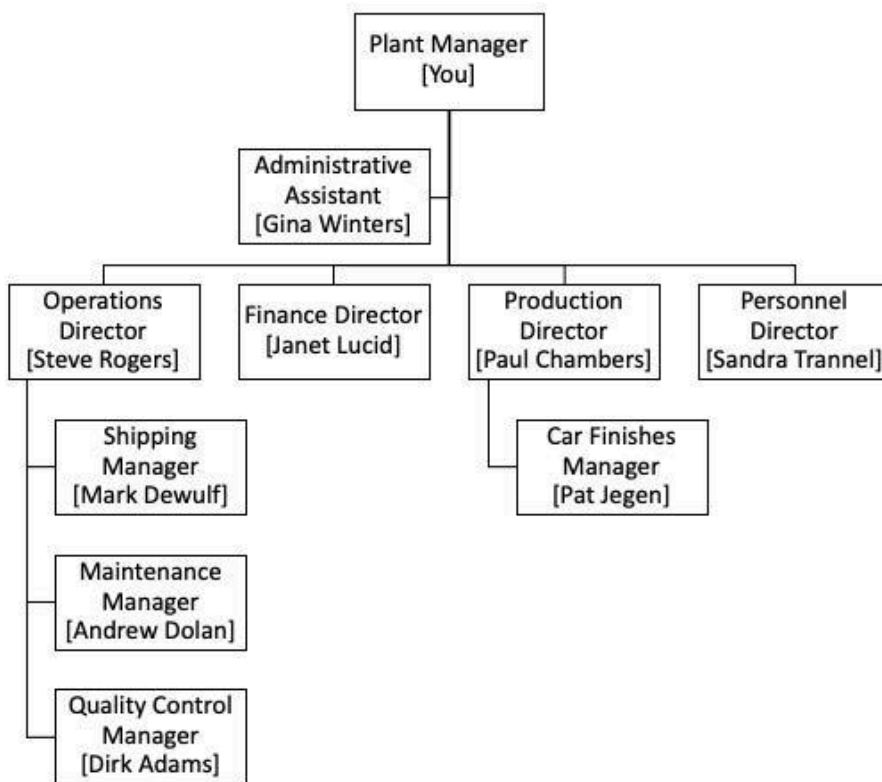
Email #01

Date: Tuesday, December 11, 2018

From: Paul Chambers, Production Director

Subject: Upcoming conference

Hi,

Last week, David Burnam arranged to meet with me on January 7, 2019. This meeting is about a new line of wind-resistant paint products to be used in airplane construction. I recognize that my contribution is useful in this phase, but I had planned since last summer to attend a conference in Santa Fe. The conference takes place February 1–3, and is completely dedicated to new production processes, including wind-resistant paint products. Would it be possible to

postpone the meeting until after the 7th? I would like to have an answer before Tuesday at noon.

Response #1: It would be difficult to reschedule the meeting because David will be coming here to attend it. Can you still cancel the flight to Santa Fe? Please wait for now – I will contact David and then let you know what he thinks.

Response #2: I think that it is more important for our organization that you attend the conference in Santa Fe. If I understand correctly, our meeting is also about the same theme of "innovation." In that case, we have everything to gain by first waiting for the conference and postponing the meeting until after your return.

Response #3: I don't think the conference in Santa Fe is that important. After all, it's mostly advertising from a few big companies. In any case, I wouldn't worry too much about this conference. I think we should go ahead with the meeting as planned.

Response #4: I understand that this conference has already been planned for a long time and that it is your first priority. I will move the meeting to February 15. David Burnam knows what our priorities are and won't have a problem with it.

Email #02

Date: Tuesday, December 11, 2018

From: Sandra Trannel, Personnel Director

Subject: Andrew Dolan, Maintenance Manager

Hi,

Yesterday, Steve Rogers, our Operations Director and Andrew's boss, told me that Andrew has missed two monthly staff meetings since his divorce in June. And last

Monday, he forgot to inform his people about some scheduling changes, which caused everyone quite a few problems.

Steve and I have already talked with Andrew. He said that he would do better, but he didn't look very serious. Maintenance is important, and we can't afford to have problems with one of the team leaders. All ideas about this are welcome.

Response #1: Maybe you can be patient for a while longer, Sandra. There is a good chance that over time, Andrew will find a way to solve his problems on his own. Maybe you can try talking with him again.

Response #2: Next week, I will speak with him personally about his performance. Make an appointment for him for Thursday morning. It is not acceptable that one person disrupts the production process by neglecting his responsibilities.

Response #3: Maybe it would be better if we excused Andrew from the monthly meetings for the next few months. I think you can decide how to best solve this. If you would like, I can also speak with him myself later this week.

Response #4: I will speak with him next week. The fact that I am getting involved will send a clear signal. If there is no improvement after our meeting, then I will take disciplinary action.

Email #03

Date: Wednesday, December 12, 2018

From: David Burnam

Subject: Trip to India

Hi,

Before I got promoted, I made plans to take a trip to India, which is an important market for us. If you go, you will visit chemical plants and industrial facilities in the New Dehli area. Your host will be Rashid Gupta; he is Assistant Director of the Indian Ministry of Foreign Affairs. The trip is planned for the end of January, which fit best into Mr. Gupta's schedule. This trip may be a chance for us to improve our competitiveness. Let Gina, your administrative assistant, know as soon as possible what you want to do. If you decide to go, she will book your trip.

Response #1: This seems like an important opportunity for the future of our company. Trips like these help us keep in close contact with our most important markets. In addition, they give our image some international appeal.

Response #2: It seems interesting to exchange information with some international businesses. Can you provide more information about your plans for this trip? I will tell Rashid Gupta that I may come in your place. I will also take a look at your plans and Gupta's plans, and then let you know what I decide.

Response #3: I don't know if we have much to learn from the chemical facilities in India. We should also be careful that we don't give away too much information, in order to prevent our newest products from being copied over there.

Response #4: Business is really busy at the moment. We are starting a new production process, so these weeks are hectic. Your proposal is attractive, but I don't know if I really have time right now for such a trip.

Email #04

Date: Wednesday, December 12, 2018

From: Steve Rogers, Operations Director

Subject: Wrong paint cans

Hi,

We have a problem with our last shipment of 5-gallon paint cans from Erlbaum Cans, one of our manufacturers.

One of Mark Dewulf's people (in the shipping department) noticed that the handles of about 1 of every 10 cans is so loose that they almost fall off when the paint can is full.

It was a fairly big shipment from Erlbaum, and this defect could cause problems for us once the cans reach the store shelves. I am looking for some quick solutions. I would really appreciate it if you had any ideas.

Response #1: No matter what, you have to confront Erlbaum about this. It is not acceptable that they deliver inferior products. Negotiate with Erlbaum about compensation for the damages caused by this delivery. Make sure that you also get a written agreement about compensation for damages in future shipments.

Response #2: Thank you for letting me know about this. For now I actually wouldn't worry too much; no defects have been reported yet. Things will probably turn out just fine with the cans.

Response #3: Thank you for bringing this to my attention. We must prevent these cans from making it to our customers. Defective products are terrible for our image. We must urgently think about how to prevent such problems from occurring in the future.

Response #4: For now, I wouldn't take any action. I'm counting on you to keep an eye on the problem and figure out where the problem originated. Let's just wait and see for now.

Email #05

Date: Thursday, December 13, 2018

From: David Burnam

Subject: Safety concerns (to all plant managers)

Patricia Ayuda, our corporate Vice-President of Human Resources, told me yesterday that James Hernandez, our CEO, approved the annual mistake-reduction program. Apparently, each year too many errors are made, and this year's results have again reached an alarmingly high level. As happens each year, each of you has to propose some realistic strategies to reduce the number of human errors occurring in your branch. Sorry, but these plans have to be on my desk on Monday.

Rate how likely it is that you would use each response option.

Response #1: Whew, this is fast. I will only be able to discuss this briefly with the production manager on Tuesday afternoon. In addition, it seems more efficient if we could first talk it over with the different plant managers tomorrow, so we can coordinate our plans. See you tomorrow!

Response #2: Maybe we can start by discussing the different possibilities. Would you have a minute to discuss this with me individually? I will soon organize a round-table with the various department heads. Once I know what their opinions are, I will propose a complete plan, OK?

Response #3: Unfortunately, I will need more time to set up a program. This issue is too delicate to decide on quickly. I will think about which strategy will be best to reduce the number of errors. See you tomorrow.

Response #4: Such a clear question deserves a clear answer. I can tell you that our procedures are based on training, clear rules, and performance monitoring. We have also recently appointed a new coordinator to oversee our procedures. The production manager and I will spend a bit of time fleshing out the plan. See you tomorrow.

Email #06

Date: Thursday, December 13, 2018

From: Janet Lucid, Finance Director

Subject: Toffler file

Hi,

Herb Mueller of Toffler in River Oaks (one of our clients) wants to re-negotiate his contract on car paint. Toffler just got an order that will increase their production by 20%. Truth is, they already have our best price, but they want an even better one.

We have done good business with Toffler in the past, but Gibson Industries in Manchester is a bigger customer and has never complained about our prices. The Toffler contract comes up next year in October, but they are hoping that their new project will be fully operational by March. Herb wants an answer before next Friday.

Welcome to Chicago!

Response #1: I suggest that we take some time to look for the best solution. I have a benchmark study that compares the discounts and suggested retail prices of our most important competition. I will get this to you as soon as possible. I will keep you informed.

Response #2: On one hand we can't lower our prices even further; on the other hand, we need to make sure we don't lose Toffler as a customer. I suggest that we don't agree to their demands quite yet. We need some time to come up with a good strategy.

Response #3: I think it's clear that we can't take much more off the price. But we need to take time to think about this; we don't want to lose Toffler as a customer. Set up a meeting for us to meet with Herb next week. Let me know when and where this will take place - my preference is for Tuesday afternoon or Wednesday morning.

Response #4: This poses a problem for us. I'm afraid that we can't offer an even lower price. Unfortunately, I am leaving today for headquarters, so I can't solve this right away. For now, don't talk to anyone about this until I get back from Detroit.

Email #07

Date: Thursday, December 13, 2018

From: Paul Chambers, Production Director

Subject: Binder for the car line

Hi,

Pat Jegen, the manager of our car finishes department, told me about a new product that is a lot less expensive than a current product that we use for our automotive paint. Pat says that the new product can reduce the durability of the paint but that it still meets the criteria set by Honda. Up to now, we have always tried to exceed these criteria for durability. I don't really know what I should do in this case.

Response #1: That is really good news, because we have been looking for this for a long time. From now on you can use the new product for the car paint. I suggest that you start production as soon as possible.

Response #2: Thanks for bringing me up to speed on these new developments. I am not sure yet; I don't want to decide too quickly about this. I am leaving tonight for a meeting in Detroit, so I can't help you just now.

Response #3: We urgently need to lower the cost of our production process in order to stay competitive in the future. A less expensive product would certainly help with this. I suggest that we carry out some tests to thoroughly test the new product.

Response #4: Thanks for telling me about this, but I would still rather continue working with the current product. I don't think that this new product has much to offer.

Email #08

Date: Thursday, December 13, 2018

From: Gina Winters, Administrative Assistant

Subject: Chinese delegation

Hi,

Just to remind you, next month a group of business people from Hong Kong will visit our company. They are considering investing in Duron, and the main office wants to make a good impression. They are coming on January 14 at 10:00 AM and will leave by 2:00 PM.

Their schedule is packed at the moment. I know that David Burnam wanted Pat Jegen (manager of our car finishes department) to have the new line of car finishes ready to demonstrate to them, but I don't know how far Pat is with that. No matter what, this is their first visit to the United States. I have arranged for Paul Chambers, our production director, to give them a tour of the plant. Is there anything else I should do?

Response #1: I will find out this week how far Pat has gotten with the new line and will let you know as soon as possible. That way, you will still have plenty of time to adjust the schedule. Can you send me the detailed visit schedule?

Response #2: I want to make a good impression on these investors, so the visit needs to go smoothly. Contact Pat again to see whether he is ready with the new line. Will you also make sure that everyone on our end arrives early? There is little room in the schedule, so everything needs to start on time.

Response #3: No, I think that everything is already planned and that the schedule is set. Wouldn't it also be a good idea to give a tour of the sales offices? See if you can find someone who can take care of that.

Response #4: I don't think that you need to arrange anything for the time being. Pat has known about this visit for a long time, so I assume that everything is taken care of. Thanks in advance for arranging everything.

Email #09

Date: Friday, December 14, 2018

From: Steve Rogers, Operations Director

Subject: Chemical pollution

Hi,

In case you haven't heard, the Chicago River has been found to be heavily polluted with chemical products, about one mile from here. The Environmental Protection Agency (EPA) has traced where the chemical products are coming from, and they have established that they originate in the drainage canal, which is located on the edge of our property.

The EPA is taking this case very seriously and wants to examine our facilities, as well as those of some other companies located nearby (including Federa and Peers Plastics). Head Inspector Collins of the EPA already called you today, but you weren't here, so I just arranged it myself. I said that Monday would be fine. Any comment?

Response #1: I want complete openness during the inspection. Let the EPA do its work. I will leave you all the previous reports from the EPA so that you can prepare a bit. Also try to gather some information from our prevention advisor. Let me know if you need any more information.

Response #2: This calls for preventive measures since it could have negative consequences for the plant. In the event that we have had something to do with this pollution, we need to know before the EPA comes to inspect us. Research the case and make a plan to prevent such problems from happening in the future.

Response #3: No matter what, we must ensure that we don't get a negative report from the EPA. Make sure that no information about this report leaks to our neighboring companies. Try to discover if there are any possible problems, even before the inspection happens. Maybe this can help us prevent something even worse from happening.

Response #4: Don't panic – it hasn't yet been proven that the pollution is coming from us. It's just as likely to be Federa and Peers Plastic's fault. In past years, we have always passed all environmental inspections. It would surprise me if we suddenly have a problem now. Just let the EPA do its work.

Email #10

Date: Friday, February 20, 2015

From: Janet Lucid, Finance Director

Subject: Gibson Industries

Hi,

The people from Gibson Industries have complained that a large part of the paint that we recently delivered to them is the wrong color. According to Dirk Adams, our manager of quality control, Gibson's buyers changed their minds about three times before they placed their final order. Dirk says that Gibson got exactly what

it ordered. The paperwork is pretty sloppy, so we really don't have any proof. I am not sure how best to handle this.

Response #1: This doesn't seem like a terrible problem. Maybe you can just talk it over with the people from Gibson Industries. We have had a good working relationship with them for years, so this can get resolved quickly.

Response #2: I am counting on you to quickly resolve this misunderstanding with Gibson Industries. I have seen that there are other files here about misunderstandings and wrong orders in the past. These will help you in your discussion with Gibson Industries. I will make sure that you get the files.

Response #3: I think we had better get to the root of this problem. It is not the first time that something like this has happened, after all. I am counting on you to not only talk with the people from Gibson Industries, but to also look into how we can prevent similar problems from occurring in the future. Such mistakes can negatively affect our image.

Response #4: I am counting on you to resolve this carefully. I would not tell Gibson Industries that something was wrong with our paperwork. They are one of our most important customers, and we really can't afford to do anything that may upset them.

Motivation incentive and post-in-basket information:

If you are found to be a good fit for the company, you will earn a $0.25 signing bonus that will be paid to your mTurk account.

Thank you for completing this task. You have done everything that you need to do for Apex to determine whether or not to hire you.

Unfortunately, you did not receive an offer for the job that you applied to. You still need a new job, as your current job will end in four weeks. You recently saw an ad for another job at Apex, similar to the one you just applied for. Imagine that you apply for that job as well.

Scoring Scheme:

Email	Response	Coordinating	Decisiveness	Information management	Problem Awareness
Email 1	1	0	-1	0	0
	2	0	0	0	+1
	3	0	0	0	-1
	4	0	+1	0	0
Email 2	1	0	-1	0	0
	2	+1	0	0	0
	3	-1	0	0	0
	4	0	+1	0	0
Email 3	1	0	0	0	+1
	2	0	0	+1	0
	3	0	0	-1	0
	4	0	0	0	-1
Email 4	1	0	+1	0	0
	2	0	0	0	-1
	3	0	0	0	+1
	4	0	-1	0	0
Email 5	1	+1	0	0	0
	2	-1	0	0	0
	3	0	-1	0	0
	4	0	+1	0	0
Email 6	1	0	0	+1	0
	2	-1	0	0	0
	3	+1	0	0	0
	4	0	0	-1	0
Email 7	1	0	+1	0	0
	2	0	-1	0	0
	3	0	0	0	+1
	4	0	0	0	-1
Email 8	1	0	0	+1	0
	2	+1	0	0	0
	3	-1	0	0	0
	4	0	0	-1	0
Email 9	1	0	0	+1	0

	2	0	0	0	+1
	3	0	0	-1	0
	4	0	0	0	-1
Email 10	1	0	0	0	-1
	2	0	0	+1	0
	3	0	0	0	+1
	4	0	0	-1	0
Score /20		(X+16)*5/8	(X+20)/2	(X+20)/2	(X+24)*5/12

Each response is scored on one of four competencies (i.e., coordinating, decisiveness, information management, problem awareness). The scoring scheme indicates whether a response should be negatively scored or positively scored. For instance, the first response option for the first mail is scored for "decisiveness." If a participant indicates that he or she would be very likely to respond like this, this results in a negative score for the competence.

When scoring responses, the following calculation was used: *very unlikely* (multiply by -2), *unlikely* (multiply by -1), *likely nor unlikely* (multiply by 0), *likely* (multiply by +1), *very likely* (multiply by +2). At the end, all scores on each competency were summed and the formula in the scoring scheme was applied to transform scores to a 20-point.

Appendix C

Demographic Questionnaire

Directions: Please provide responses to the following items.

1. What is your sex? <Male/Female/Other (Please identify)>

2. What is your age? <open-ended>

3. Please identify your ethnicity. Please check all that apply. <Caucasian, Black or African American, Hispanic or Latino/Latina, Asian or Pacific Islander, Native American, Other (Please describe)>

4. Please indicate the number of hours you work during an average week. (1: Not employed; 2: Work fewer than 20 hours per week; 3: Work between 11 and 40 hours per week; 4: Work between 41 and 60 hours per week; 5: Work more than 60 hours per week.)>

5. Please select the option that best describes your employment status. <Unemployed; Hourly worker; Salaried worker; Student; Retired; Other (Please describe)>

Appendix D

Experimental Manipulations

Condition 1: Strong Situation

For the duration of this study, please imagine that you have had a 15-year career in a professional field, such as marketing. You like your career, but for a variety of reasons, you need to find a new job with a new company. You have applied for a job with Apex, a company in your area that you do not know much about. Your current employment will end in four weeks, and you really need this job.

Imagine that you have done some additional research on Apex, and you have learned a few things:

- Information about work-related responsibilities is explained to all employees very clearly.

- Managers are consistent in their expectations; they stay the course so that the company can remain reliable and reliably respond to customer's needs.

- Managers provide employees with a clear structure for how they are to carry out their work responsibilities.

- Any errors made on the job have serious consequences for the company; errors cannot be corrected easily.

Condition 2: Weak Situation

For the duration of this study, please imagine that you have had a 15-year career in a professional field, such as marketing. You like your career, but for a variety of reasons, you need to find a new job with a new company. You have

applied for a job with Apex, a company in your area that you do not know much about. Your current employment will end in four weeks, and you really need this job.

Imagine that you have done some additional research on Apex, and you have learned a few things:

- Information about work-related responsibilities is explained to all employees only minimally.

- Managers are sometimes inconsistent in their expectations; they sometimes change course so that the company can remain nimble and respond to customers' needs.

- Managers allow employees a to have a great deal of leeway in how they carry out their work responsibilities.

- Any errors made on the have few consequences for the company; errors can be corrected easily.

Manipulation Check

Please respond to the following questions about Apex, the company to which you have applied. Please note, if your answers are inaccurate, you may not receive payment for completing this study.

1. Managers provide a great deal of clarity regarding employees' work responsibilities. In other words, work tasks are clearly explained. [True/False]

2. Managers' expectations may be inconsistent. In other words, they may sometimes change directions with little notice. [True/False]

3. Employees are expected to follow standardized operating procedures. In other words, employees do NOT have a great deal of freedom in how they go about their work. [True/False]

4. Any errors made on the job are problematic, as they may result in major negative consequences. [True/False]

The remaining pages in this study will simulate an employee hiring

process. Please remember to imagine that you are the job applicant who was just

described: You are highly interested in getting a job at Apex.

Appendix E

Attention Check

How attracted someone is to an organization can have an impact on whether that person decides to drop out of the hiring process. As part of this research, we are interested in knowing more about you. Specifically, we are interested in whether you take the time to read the directions for each task. To demonstrate that you have read the directions, please ignore the questions below. Instead select *Strongly Disagree* for each item.

[A 5-point scale was provided for each item.]

- If I was a job applicant, this company would be a good place to work.

- If I was a job applicant, I would not be interested in this company except as a last resort.

- If I was a job applicant, this company would be attractive to me as a place of employment.

- If I was a job applicant, I would be interested in learning more about this company.

- If I was a job applicant, a job with this company would be very appealing to me.

Appendix F

Curve Estimates of Significant Nonlinear Facet–Criterion Relationships

Duty when predicting coordination: significant quadradic relationship
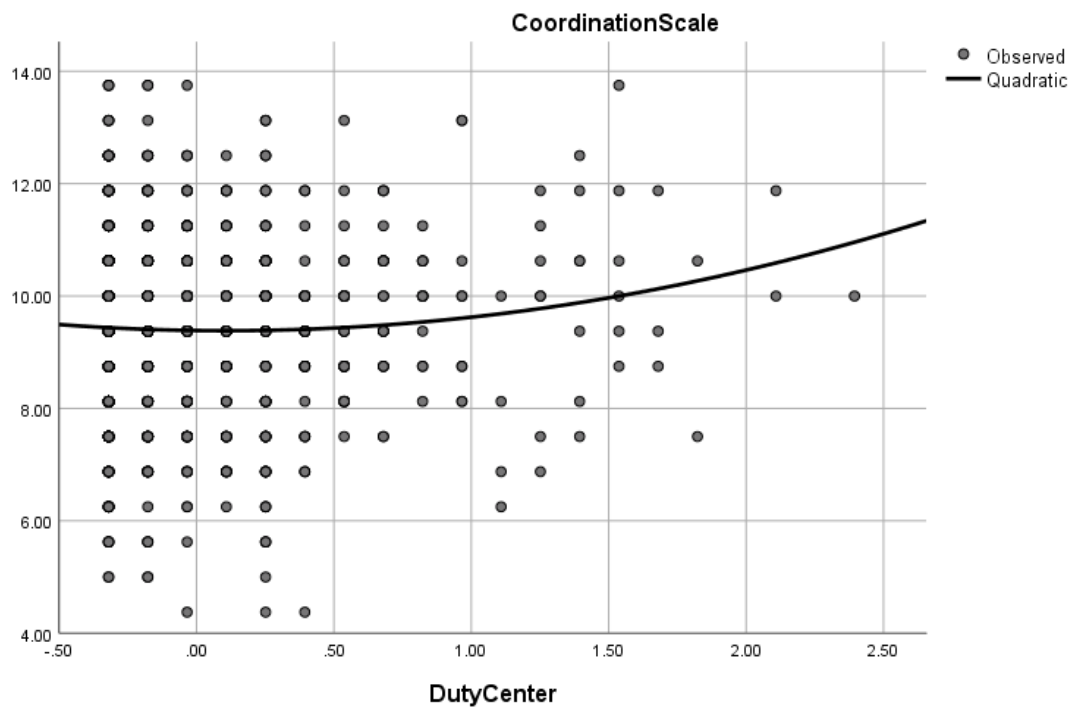
Achievement when predicting coordination: significant cubic relationship
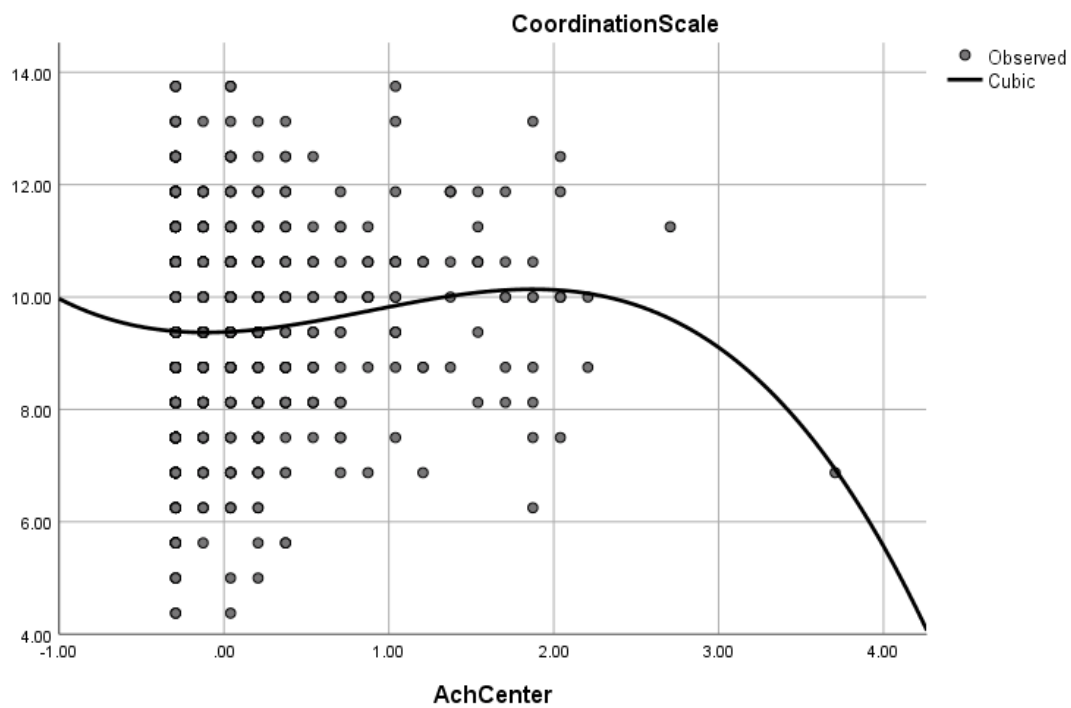
Discipline when predicting coordination: significant cubic relationship
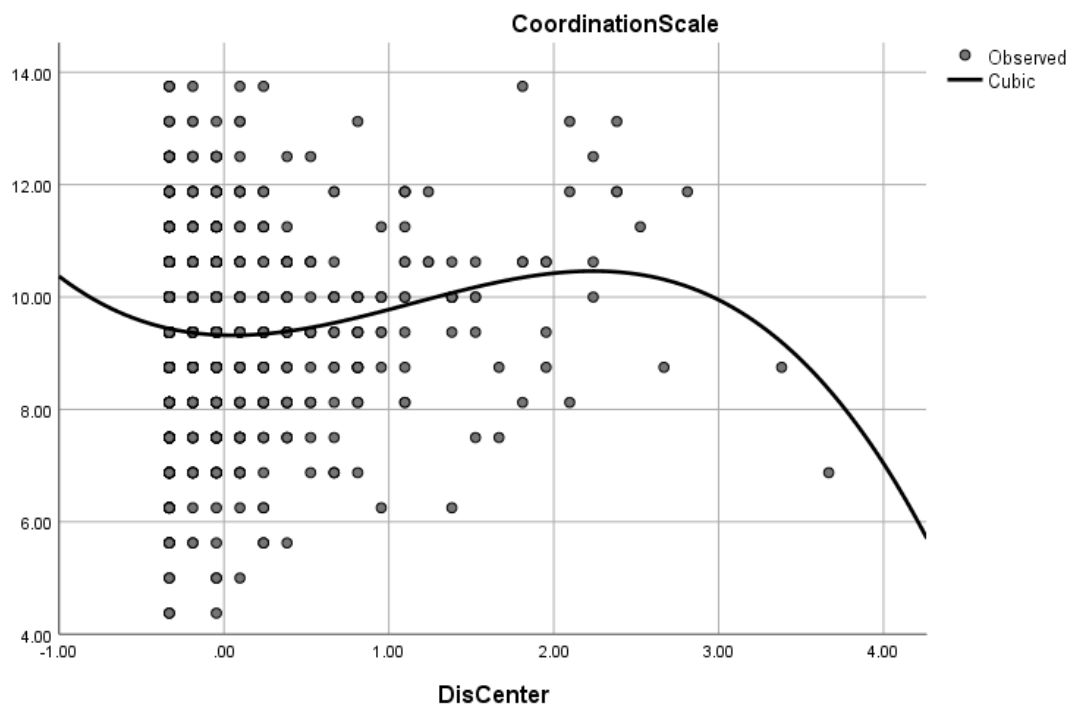
Cautiousness when predicting coordination: significant quadradic relationship
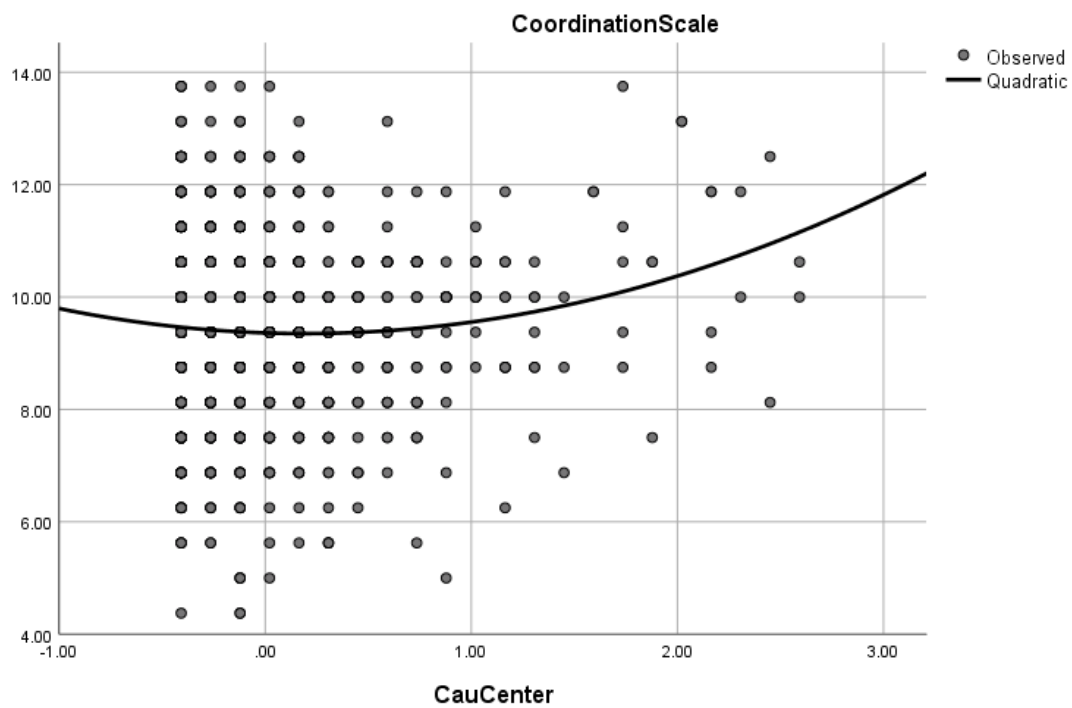
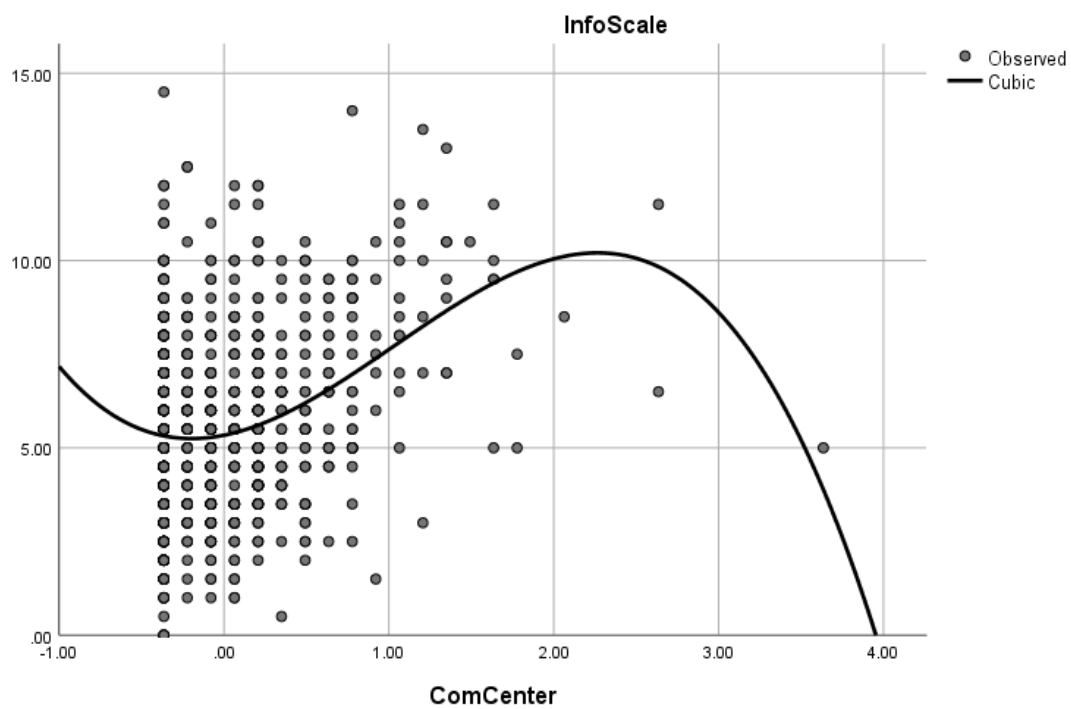Competence when predicting decisiveness: significant cubic relationship
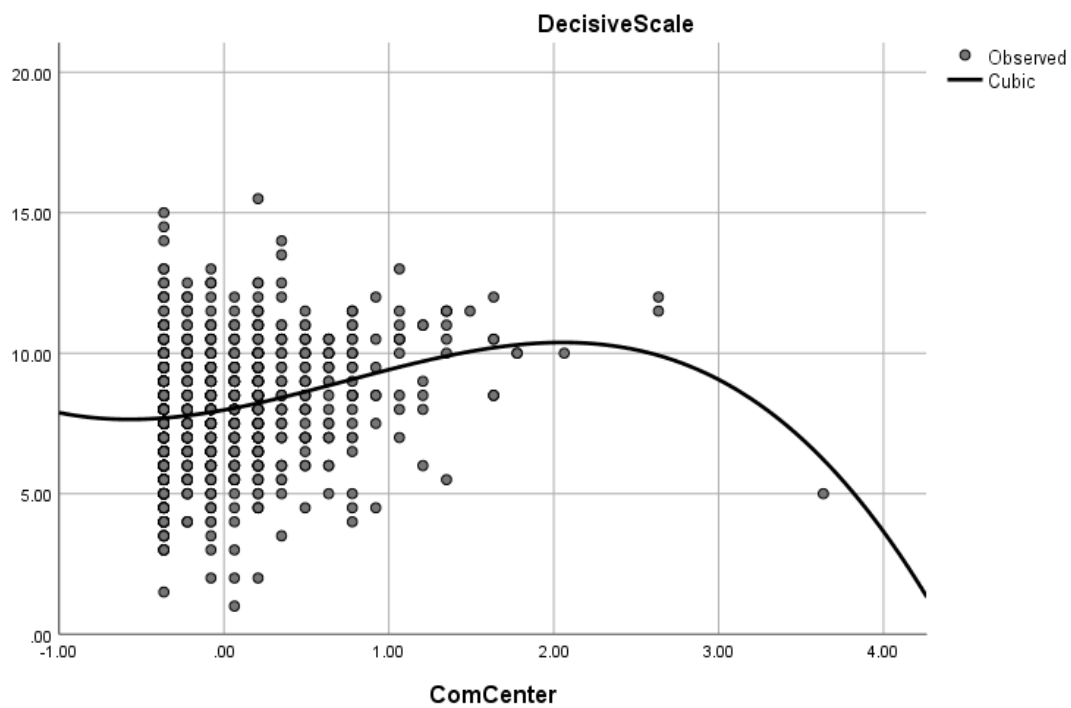
Achievement when predicting decisiveness: significant quadradic relationship
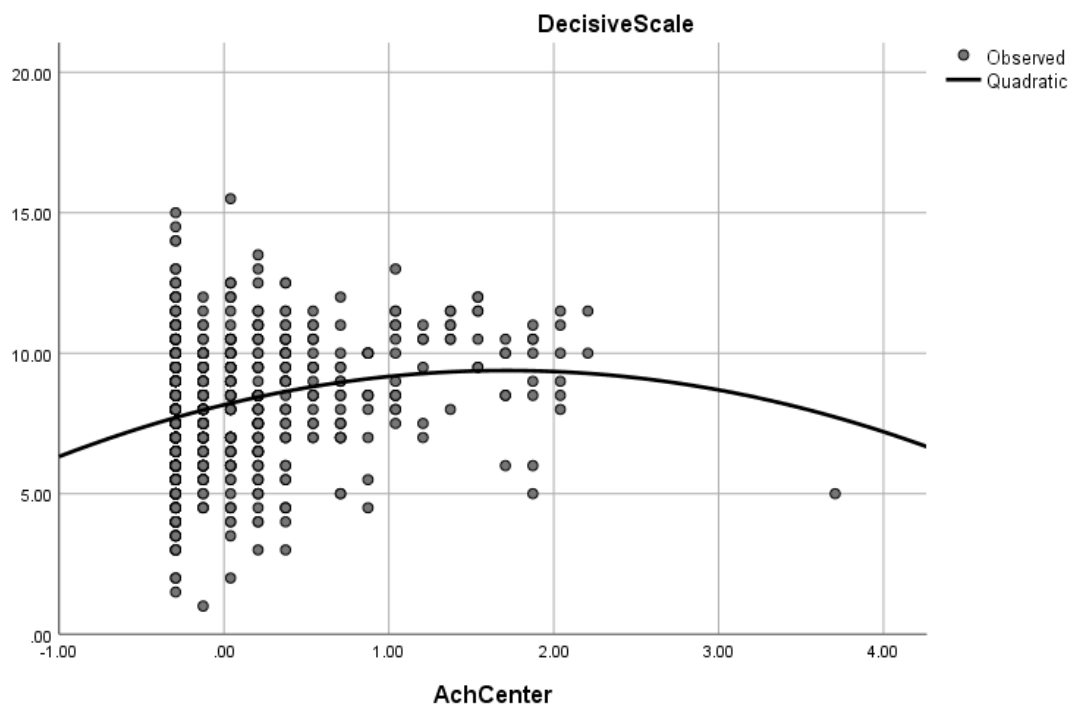
Discipline when predicting decisiveness: significant quadradic relationship
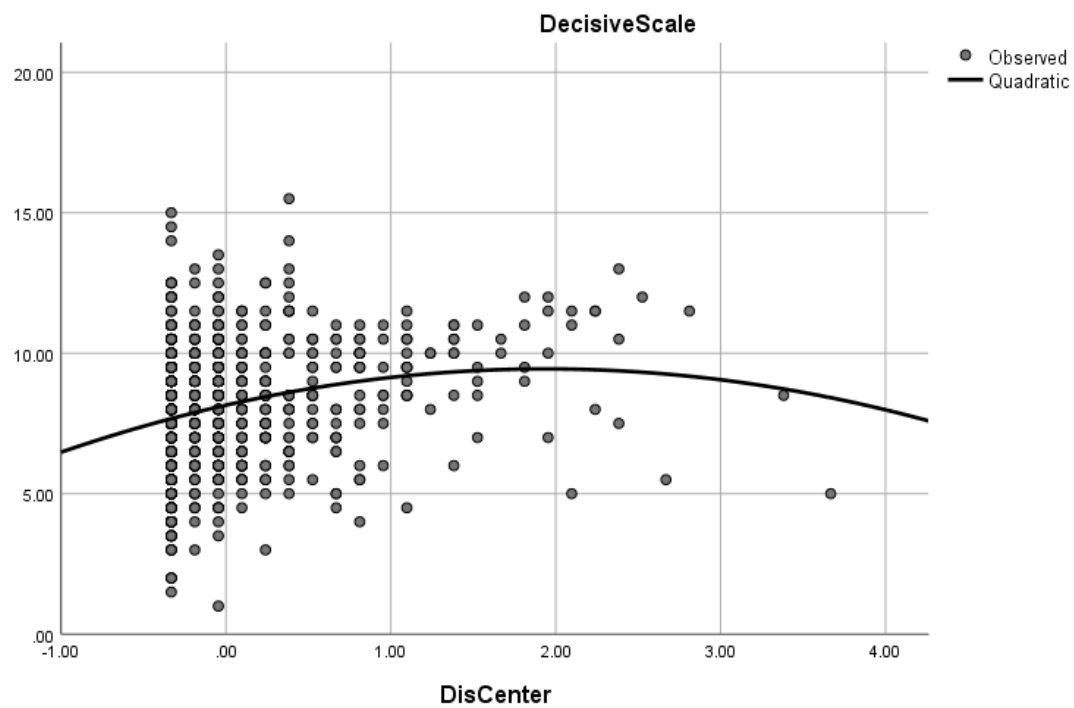
Competence when predicting information management: significant cubic relationship

Duty when predicting information management: significant quadradic relationship

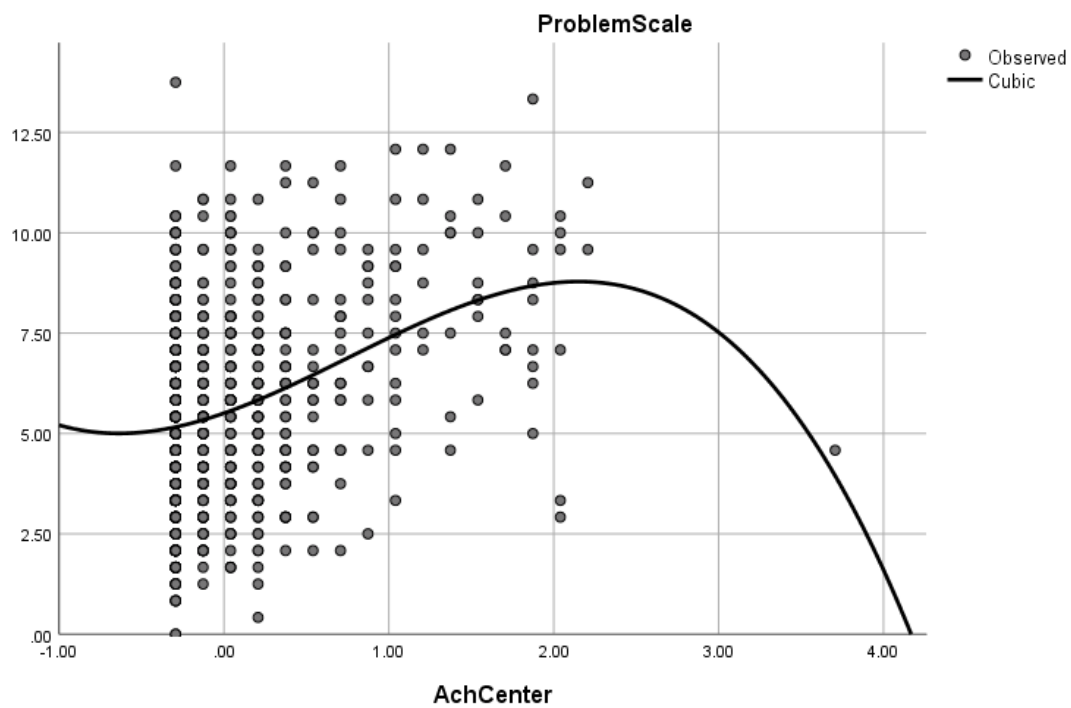Discipline when predicting information management: significant cubic
relationship

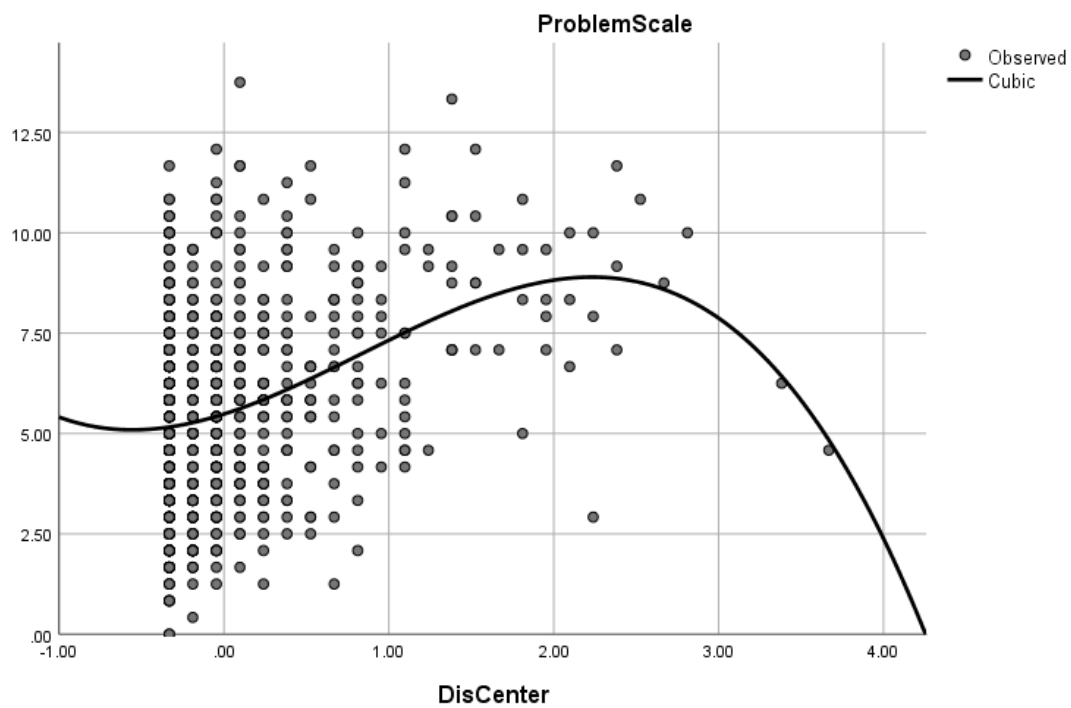Competence when predicting problem awareness: significant cubic relationship

Duty when predicting problem awareness: significant quadradic relationship

Achievement when predicting problem awareness: significant cubic relationship

Discipline when predicting problem awareness: significant cubic relationship

Cautiousness when predicting problem awareness: significant cubic relationship