DISSERTATION

COGNITIVE ABILITY TESTING FOR EMPLOYEE SELECTION:

IMPLICATIONS FOR AGE DISCRIMINATION

Submitted by

Megan N. Naude

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2018

Doctoral Committee:

Advisor: Gwen Fisher

Alyssa Gibbons
Matthew Rhodes
Chris Henle

ABSTRACT

COGNITIVE ABILITY TESTING FOR EMPLOYEE SELECTION: IMPLICATIONS FOR AGE DISCRIMINATION

Existing theory and empirical research suggest that tests of fluid cognitive abilities have the potential to lead to age-based adverse impact and may be stronger predictors of job performance for younger job candidates compared to older job candidates. However, the evidence suggests that tests of crystallized cognitive abilities are not as susceptible to age-based adverse impact issues and should be strong predictors of job performance for candidates of any age. The two present studies used cognitive ability test scores collected from management employees in a large company in the United States in conjunction with supervisory performance ratings to examine adverse impact based on age, linear relations of test scores with age, and differential validity and prediction based on age. In the first study, a sample of N = 214 employees completed a test of fluid cognitive abilities, and in the second study, a sample of N = 232 employees completed a test of crystallized cognitive abilities. Contrary to hypotheses, results indicated that age-based adverse impact was more likely to be present for the test of crystallized abilities, age was negatively related to test performance for both tests, and neither test resulted in significant differential validity or prediction for the two age subgroups. Implications for theory and practice are discussed.

ACKNOWLEDGEMENTS

I am beyond grateful for the guidance and support of many people throughout my graduate school journey. First, I would like to thank Gwen Fisher for her dedication and encouragement over the last five years. I am also thankful to my dissertation committee—Alyssa Gibbons, Chris Henle, and Matt Rhodes—for their insightful feedback throughout both the empirical comp and dissertation processes.

Without a doubt, I would not have made it to this point without the love and friendship of my fellow CSU I/O grad students. I feel incredibly lucky to have had the chance to work alongside my best friends. I am particularly grateful to my cohort, both official and unofficial members, for being there through all the laughter and tears (thanks to you all, it was mostly laughter with only a few tears). The only hard part about being done with grad school is the fact that I had to leave my Fort Collins family behind.

I cannot begin to express my appreciation for my family—Mom, Dad, Lauren, and Ashleigh—for their love and support. My motivation to reach this achievement has stemmed both from having my parents as role models as well as aspiring to be a role model for my sisters.

Finally, thank you to Ryan for continuing to make me laugh during the toughest part of this journey. There’s no one else I’d rather have beside me as I finally experience this mysterious yet magical concept of life post-dissertation.

TABLE OF CONTENTS

Introduction

The topic of age discrimination in the workplace continues to be relevant in current times, especially given the aging workforce and changing norms regarding career patterns. The proportion of workers aged 55 and older in the workforce has steadily increased and is projected to continue increasing (U.S. Bureau of Labor Statistics, 2015). Additionally, those in the workforce, including older workers, are less likely to remain with one employer throughout their careers (Wang, Olson, & Shultz, 2012), and job and organizational mobility continue to become more prevalent across generations (Lyons, Schweitzer, & Ng, 2015). In the 1970s, Americans held an average of 7 jobs (defined as an uninterrupted period of work with a particular employer) during their careers (Kolb, 1984). Findings from the National Longitudinal Survey of Youth, conducted from 1979 to 2015, showed that individuals born between the years of 1957 and 1964 held an average of 11.9 jobs between ages 18 and 50. Although the majority of these jobs were held before age 34, participants in the study held an average of 2.9 jobs between ages 35 and 44 and 1.7 jobs between the ages of 45 and 50. Additionally, among workers starting new jobs between the ages of 35 and 44, 36 percent left the job in less than a year and 75 percent left the job in fewer than 5 years (U.S. Bureau of Labor Statistics, 2017). Economic recession may also raise the average age of job seekers. Wanberg, Kanfer, Hamann, and Zhang (2016) highlighted the impact of unemployment on older workers and presented meta-analytic evidence of a negative relationship between age and reemployment status.

Given the changing demographics of the aging workforce, it has become imperative that organizations consider issues related to older workers (Truxillo, Cadiz, & Rineer, 2014). From a business perspective, organizations may specifically seek older workers with more experience to meet business needs. Many companies choose to hire experienced employees because they possess critical knowledge, skills, and abilities that meet business needs and can bring diverse experience and knowledge into organizations (Almeida, Dokko, & Rosenkopf, 2003; Rao & Drazin, 2002). Indeed, in a survey of staffing professionals, Rynes, Orlitzky, and Bretz (1997) found that organizations that strategically shifted toward experienced hiring as opposed to new graduate hiring did so because of an increase in demands of the job and necessary skills or because jobs required more immediate productivity from new hires. The results of the study also suggested that a majority of positions across organizations were filled through experienced hiring rather than new graduate hiring. Thus, age discrimination at the selection stage is a concern for organizations, given that applicant pools may include candidates ranging from recent college graduates to those in the later stages of their careers (Fisher, Truxillo, Finkelstein, & Wallace, 2017).

In a previous study (Naude, 2018), I examined the extent to which the amount of time between college graduation and taking a cognitive ability test was related to test performance. I examined linear relations between time since graduation and test performance and also applied item response theory methods to test for the presence of differential item functioning as a function of the amount of time that had passed between college graduation and the test date. Although this previous study made an important contribution to the literature by calling attention to the lack of research on issues relevant to experienced job candidates, the results of the study did not identify any significant relations between time since graduation and cognitive ability test scores. However, examination of the descriptive statistics for the sample included suggested that the sample was not ideal for testing these hypotheses, given that the distribution of the time since graduation variable was heavily skewed such that a large majority of participants were still in school or had recently graduated when taking the cognitive ability test.

The purpose of the current studies was to explore age discrimination in the context of selection, and more specifically cognitive ability testing for employee selection purposes. These studies extend the prior work of Naude (2018) in several ways. First, whereas the previous study focused on the effects of time since graduation, the present studies focused on age-related differences in cognitive ability test performance and included samples for which age data was available for all participants. The present studies also address the heavily skewed age distribution of the previous by using samples for which age was more normally distributed. The present focus on age also means that the practical implications are greater, given that those 40 years of age and older are a legally protected group, and age discrimination lawsuits can be costly to organizations in terms of time, money, and reputation. Second, the previous study only examined scores on tests of fluid cognitive abilities; Study 1 examined fluid cognitive abilities from an age perspective and Study 2 examined age-related differences in crystallized ability test scores. Third, the present studies used job performance data to test theoretically-driven hypotheses regarding differential validity and differential prediction, which were not addressed in the previous study.

The following sections I first discuss the implications of age discrimination from both legal and talent pipeline perspectives. Next, I review the existing literature related to cognitive ability, aging workers, selection testing, and the intersection of these topics. I also identify the current gaps in the literature and explain how my studies addressed these issues. Finally, I present the details of two studies using organizational data to address these gaps.

Age Discrimination in the Workplace

Legal issues. Organizations in the United States must always be aware of the potential legal consequences of any decisions based on selection tests. Title VII of the Civil Rights Act

(1964) prohibits employment discrimination based on race, color, religion, sex, or national origin. Age was not established as a protected class under Title VII, largely because age discrimination was perceived to be a consequence of stereotyping rather than intentional actions (Gutman, Koppes, & Vodanovich, 2010). However, the Age Discrimination in Employment Act enacted in 1967 prohibits employment discrimination in the U.S. on the basis of age, originally specifying individuals between 40 and 65 years of age as a protected class. It was later amended to raise the age limit to 70 years of age and again in 1986 to remove age limits altogether, resulting in protection for workers age 40 and older (Neumark, 2009). Whereas discrimination claims within Title VII are based on differential treatment between members of a protected class and those who are not part of a protected class (e.g., women being treated less favorably than men; minorities being treated less favorably than non-minorities), protected class under the ADEA is treated as continuous rather than categorical (Gutman et al., 2010). In the case of *O'Connor v. Consolidated Coin Caterers* (1996), the Supreme Court ruled that age discrimination is determined based on whether "a replacement is substantially younger than the plaintiff." Thus, a valid case for age discrimination can be made when, for example, a 45 year-old replaces a 55 year-old employee (Gutman et al., 2010). Although it was not initially clear, Supreme Court rulings in cases such as *Smith v. City of Jackson* (2005) have confirmed that adverse impact claims can be made under the ADEA (Gutman et al., 2010).

Two U.S. Supreme Court cases that have been impactful regarding adverse impact based on employee selection procedures: *Griggs v. Duke Power Co*. (1971) and *Albemarle Paper Co. v. Moody* (1975). *Griggs v. Duke Power Co*. originated from 13 African American employees claiming a violation of Title VII due to Duke Power's requirement that employees wishing to work in any department other than the Labor Department either have a high school diploma or pass two aptitude tests. According to the Supreme Court, African American candidates were less likely to meet these requirements for selection because they had “long received inferior education in segregated schools.” The Court also questioned the legitimate business necessity of a high school diploma for the jobs in question and ultimately pointed to the guidelines established by the Equal Employment Opportunity Commission (EEOC) regarding the job-relatedness of selection tests. Thus, the case reinforced the stipulation that organizations be able to show evidence that selection tests are significantly related to job performance in order to legally defend the tests against claims of discrimination (Outtz, 2011).

In the case of *Albemarle Paper Co. v. Moody*, claims were once again made that selection tests used by the company discriminated against African American job candidates. A district court initially ruled that the company had shown evidence of job-relatedness through a validation study and was therefore not liable for discrimination. However, the United States Court of Appeals reversed the ruling, pointing to issues with the validation study conducted by Albemarle. Specifically, the company had conducted the validation study by showing a relationship between test scores and only one of the several job groups it was used for, using supervisor rankings as the criteria without providing any specific job performance criteria, and using only experienced White employees as subjects. The company appealed the ruling to the Supreme Court which upheld the decision, reiterating the deficiencies of the validation study. The Court also stated that “even if the employer meets its burden of showing validity, the plaintiff has the opportunity to show that other tests or selection devices with similar adverse effects would serve the employer’s interest just as well.” Thus, selection tests with similar predictive validity but resulting in less adverse impact are always preferable (Outtz, 2011).

Established in 1978 by the EEOC, the *Uniform Guidelines on Employee Section Procedures* (*Guidelines*) provided the 80% or four-fifths rule for determining whether adverse impact, defined as a “substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group” (Sec. 4D) has occurred. In the case of selection testing, if the rate at which a subgroup of candidates (based on race, sex, ethnic group, etc.) passes a selection test is less than 80% of the rate at which another group passes the same test, evidence of adverse impact exists (Guion, 2011). The *Guidelines* also made it clear that organizations can legally defend tests that result in adverse impact by showing evidence of the predictive validity of the test but that alternative tests with equal predictive validity and less adverse impact should be used (Outtz, 2011).

Thus, organizations using selection assessments that systematically disadvantage protected classes of job candidates are vulnerable to claims of adverse impact. Organizations may experience negative outcomes even if they win an adverse impact case by providing evidence of the relation between performance on the test in question and performance on the job. For example, the time and legal expenses associated with litigation and the potential to earn a negative reputation from such accusations in addition to ethical considerations provide motivation to ensure that selection tests are used in a fair manner (Cascio & Boudreau, 2011). In 2017, over 18,000 age discrimination charges were filed with the EEOC, and 13% of those were upheld or settled resulting in $90.1 million in monetary benefits (EEOC, 2017).

Benefits of age diversity. In addition to legal concerns, the workforce diversity literature provides another compelling argument for organizations to avoid discrimination and maximize diversity. More specifically, Konrad (2003) posited that the business case for diversity stems from three arguments. First, a more diverse labor force increases the need to recruit top talent from a wide variety of demographic categories. Second, increases in the diversity of U.S. society and the globalization of the marketplace means that organizations with more diverse employees will have an advantage when selling products and services to a range of consumers. Finally, Konrad (2003) claimed that demographically diverse groups are better at problem-solving and creativity tasks due to the range of information, experience, perspectives, and cognitive styles of group members. Specific to age, the findings of several studies provide evidence suggesting that age diversity has positive implications at both the employee and organizational levels. For example, Kunze, Boehm, and Bruch (2011) proposed and found support for a model in which age diversity was positively associated with firm performance through the negative influence on perceptions of age discrimination climate and collective affective commitment. In other words, when organizations were characterized by more age diversity, employees were more likely to perceive that the company did not discriminate based on age and, in turn, were more likely to experience affective commitment, ultimately leading to better organizational performance. Similarly, in a study of 93 organizations and over 14,000 employees, Boehm, Kunze, and Bruch (2014) found that age-diversity climate was positively related to firm performance and negatively related to turnover intentions.

Age discrimination as a barrier for experienced hiring. Organizations often target experienced job candidates as a strategy to bring in knowledge and skills that businesses need (Rynes et al., 1997). Many organizations expect that more experienced job candidates will be able to perform at their full potential at a faster rate compared to candidates with limited or no work experience (Dokko, Wilk, & Rothbard, 2009; Rynes et al., 1997). Although initial empirical findings regarding the relationship between prior work experience and job performance were mixed (McDaniel, Schmidt, & Hunter, 1988), Dokko and colleagues (2009) found support for a mediated model in which prior related job experience predicted job performance. This occurred through an increase in task-relevant knowledge and skill and suppressed a direct negative effect of experience on performance due to "baggage" from previous experience that employees bring with them to new organizations. More recently, Uppal, Mishra, and Vohra (2014) found support for a positive relationship between previous related work experience and job performance as well as moderating effects of the Big Five personality traits.

Other studies investigating the effects of job experience have also found positive relationships with self-efficacy (Morrison & Brantner, 1992), newcomer adjustment in teams (Beus, Jarrett, Taylor, & Wiese, 2014), and response time to work related challenges (Beyer & Hannah, 2002), and negative relationships with depression and hostility (Motowidlo, Packard, & Manning, 1986) and work-family conflict (Byron, 2005). Thus, age discrimination leads organizations to forgo the knowledge and skills that older and more experienced workers bring (Tillsely & Taylor, 2001). However, it should also be noted that several studies have found negative relations between years of job experience and performance for certain professions. For example, Choudry, Fletcher, and Soumerai (2005) reviewed studies related to years of medical practice and outcomes including medical knowledge and health care quality and found that the majority of studies reported negative relations between experience and performance outcomes. Additionally, Tracey, Wampold, Lichtenberg, and Goodyear (2014) found negligible relations between years of experience and the professional skills of psychotherapists. Thus, although job experience generally predicts more positive work outcomes, relationships between experience and performance are not unitary across domains.

Sources of Age Discrimination in Selection

In order to address age discrimination in the workplace, it is important to first understand the ways in which it occurs. Although age discrimination can be manifested in many ways including lower ratings of performance (Posthuma & Campion, 2009) and more severe consequences of performance problems (Rupp, Vodanovich, & Credé, 2006), this study focuses on age discrimination at the employee selection stage.

Subjective biases. When hiring decisions are made based on subjective judgments, stereotypes can lead to discriminatory selection practices that disadvantage older candidates. Finkelstein and Farrell (2012) explained the tripartite view of age bias in the workplace. The cognitive component, stereotyping, stems from existing beliefs about older job candidates based solely on their association with a larger group of older adults. Ng and Feldman (2012) focused on six of the common stereotypes about older workers—that they are less motivated, less willing to participate in training and career development, more resistant and less willing to change, less trusting, less healthy, and more vulnerable to work-family imbalance. They provided meta-analytic evidence to support the stereotype regarding willingness to participate in training and career development activities, but the other five stereotypes were not supported by the data. Nevertheless, age stereotypes continue to be a source of age discrimination in the workplace (Fisher et al., 2017).

The affective component of age bias relates to individuals' attitudes toward older workers. Attitudes stem from the stereotypes associated with older workers or older adults in general and therefore can be both positive and negative (North & Fiske, 2015). For example, positive attitudes toward older adults are generally associated with stereotypes regarding increased warmth and kindness (Cuddy & Fiske, 2002; Kite, Stockdale, Whitley, & Johnson, 2005). However, the literature suggests that general attitudes toward older workers are more negative, relating to stereotypes of illness, irrelevance, and incompetence (North & Fiske, 2012). In fact, Cuddy, Norton, and Fiske (2005) found that people generally viewed older adults as incompetent, even when their behavior did not align with this perception.

Finally, the behavioral component of age bias relates to the tendency for individuals to treat older workers or job candidates in a certain way because of their association with a social subgroup (i.e., older adults). This differential treatment ultimately leads to age discrimination and can manifest in a number of ways (Finkelstein & Farrell, 2012). Bal, Reiss, Rudolph, and Baltes (2011) conducted a meta-analysis to examine the influence of age on various workplace outcomes and found moderate negative effects of age on career advancement (r = -.21), selection (r = -.30), and performance evaluations (r = -.24). Although age discrimination in the workplace can occur as a result of individuals' subjective judgments and consequent decisions regarding outcomes such as selection, promotion, and performance appraisal, discrimination can also occur in association with objective tests that do not rely on any subjective judgment (Fisher et al., 2017).

Cognitive ability testing. The specific focus of this study is age discrimination as a result of using tests of cognitive ability in the selection process, and specifically in multiple-hurdle selection systems. This is an important distinction because in multiple-hurdle selection systems only those candidates who meet the minimum criterion or cutoff score on an initial selection assessment can progress to the next stage of the selection process (Cascio & Aguinis, 2010). For example, if a candidate scored below the cutoff score on an initial cognitive ability test, he or she would not be invited to complete an interview. Thus, when selection assessments with automated scoring procedures and predetermined cutoff scores are used early in the selection process, age discrimination can occur before subjective biases due to stereotypes can be introduced.

Cognitive Ability Theory

The Cattell-Horn-Carroll (CHC) model of cognitive ability represents the amalgamation of the two most prominent theories of cognitive ability—the Cattell-Horn *Gf-Gc* model (Cattell, 1943; Horn, 1968, 1985; Horn & Cattell, 1966) and the Carroll Three-Stratum model (Carroll, 1993). The CHC model is widely accepted as the most empirically supported theory of cognitive abilities (Alfonso, Flanagan, & Radwan, 2007; McGrew, 2009; Wee, Newman, & Joseph, 2014) and guides the development and interpretation of many assessments of cognitive abilities (Flanagan, McGrew, & Ortiz, 2000). The highest-order factor of the CHC model is general ability, or *g*, and represents common variance measured across all types of cognitive ability tests. Second-order factors represent broad abilities or variance common to a subset of cognitive ability tests and third-order factors represent narrow abilities or common variance across an even smaller subset of tests (Wee et al., 2014). For example, the broad ability of knowledge of reading and writing, (i.e., *Grw*) can be measured by vocabulary tests, reading comprehension tests, analogy tests, etc. The narrow ability of spelling falls under *Grw* and is measured specifically by spelling tests.

The CHC model originally consisted of 10 broad abilities and more than 70 narrow abilities (Salgado, 2017), although extensions to the model continue to be explored, and as many as 16 broad abilities and many more narrow abilities have been proposed (Schneider & McGrew, 2012). Two of the most commonly measured broad cognitive abilities are fluid abilities (i.e., *Gf*) and crystallized abilities (i.e., *Gc;* also referred to as comprehension-knowledge abilities). Fluid abilities can be defined as "…the use of deliberate and controlled mental operations to solve novel problems that cannot be performed automatically" (McGrew, 2009, Table 1), and include narrow abilities of deductive and inductive reasoning as well as quantitative reasoning. Crystallized abilities are defined as "a person's breadth and depth of acquired knowledge of the language, information and concepts of a specific culture, and/or the application of this knowledge" (McGrew, 2009, Table 1). Although some researchers have recognized domain specific knowledge (i.e., *Gkn*) as a broad cognitive ability distinct from *Gc* (e.g., McGrew, 2009; Schneider & McGrew, 2012), job-specific knowledge is generally thought to be an indicator of *Gc* (Salgado, 2017). Further, Stanek and Ones (2017) recently proposed a modification to the CHC model such that Acquired Knowledge is added as a higher order construct comprised of three preexisting broad abilities—quantitative ability, verbal ability, and domain specific knowledge. For the sake of clarity, in this study job-specific knowledge is considered to be part of *Gc*.

Cognitive Ability Testing in Selection

Brief overview of selection testing. Personnel selection tests are a crucial part of the hiring process for organizations and job candidates and are used to gather more information about the knowledge, skills, abilities, and other characteristics of job candidates to predict their future performance on the job. Guion (2011) defined a test as "an objective and standardized procedure for measuring a psychological construct using a sample of behavior" (p. 485). The use of tests for personnel selection can be traced back to the early 1900s (Vinchur & Koppes, 2012), although many of the initial popular assessment methods such as graphology (the analysis of handwriting) and phrenology (the analysis of the shape of one's head) have since been deemed unscientific and invalid ways of selecting employees (Chamorro-Premuzic & Furnham, 2010). Today, common types of selection tests include a variety of different methods, such as behavioral interviews consisting of questions about a candidate's previous behaviors related to the job; knowledge or ability tests; personality inventories; work samples that allow candidates to provide examples of relevant work they have completed in the past; and assessment centers in which candidates engage in activities relevant to the job for which they applied (Cascio & Aguinis, 2010). Carless (2009) reported evidence that the prevalence of selection testing has increased in recent years and presented several reasons for this, including the increase of internet testing that has reduced the costs and increased ease of testing and a growing awareness among practitioners of testing as a best practice for selection. From a practical standpoint, cognitive ability assessments have significantly higher utility than many other selection methods, given that they are generally more cost- and time-effective (Dilchert, 2017).

In their recent review of 100 years of research findings related to selection methods, Schmidt, Oh, and Shaffer (2016) posited that the predictive validity of a selection test is the most important property from a practical standpoint, as an increase in predictive validity translates to an increase in the economic utility of a test. However, given the legal and practical implications of discrimination discussed in the previous section, the extent to which a test results in subgroup differences is also a concern for practitioners. For these reasons, the use of cognitive ability tests in particular has been one of the most widely debated issues in applied psychology (Murphy, Cronin, & Tam, 2003; Ployhart, 2006).

Predictive validity of cognitive ability tests.

Theoretical evidence. The underlying theoretical explanation for the positive relation between cognitive ability and job performance is that general cognitive ability assesses an individual's potential for learning (Sternberg & Detterman, 1986; Ones, Dilchert, & Viswesvaran, 2012). Specifically, cognitive ability is both directly related to job performance and indirectly related to job performance through job knowledge (Borman, White, Pulakos, & Oppler, 1991; Hunter, 1986; McCloy, Campbell, & Cudeck, 1994; Schmidt, Hunter, & Outerbridge, 1986). In other words, cognitive ability is positively related to both the amount of job knowledge acquired and the speed at which it is acquired, and higher job knowledge translates to higher job performance (Schmidt, 2002). The increased job knowledge associated with higher cognitive ability can manifest as both declarative knowledge, consisting of facts, rules, and procedures, as well as procedural knowledge, representing knowledge of how to carry out a specific task (McCloy et al., 1994; Ones et al., 2012). The well-documented positive relation between cognitive ability and training performance (Hülsheger, Maier, & Stumpp, 2007; Hunter, 1983; Hunter, 1985; Salgado & Anderson, 2002; Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003) is further evidence of the association between cognitive ability and acquired job knowledge (Ones et al., 2012). Despite the amount of theoretical background related to cognitive ability and job performance, however, Hunter (1989) argued that "...the fact that general cognitive ability predicts job performance on all jobs need not be theoretically proven. It can be demonstrated by brute force empirical studies…The theoretical basis for validity is shown in the data that relates ability, knowledge, and performance" (p. 11).

Empirical evidence. Evidence from thousands of studies has consistently demonstrated that cognitive ability is the best available predictor of job performance (Hunter, 1986; Hunter & Schmidt, 1996; Ree & Earles, 1992; Schmidt, 2002; Schmidt & Hunter, 1981; Schmidt, Shaffer, & Oh, 2008; Schmidt et al., 2016). Accordingly, cognitive ability tests are one of the most prevalent selection procedures used in organizations (Salgado, 2017). A 2014 study conducted by SHL with over 1,400 international companies found that 59% of companies used cognitive ability tests, with 47% using cognitive ability tests for selection purposes (Krantowitz, 2014). A more recent global study conducted by Mercer found that 56% of companies surveyed were already using cognitive ability assessments to select talent, and 30% planned to start using cognitive ability assessments in 2017 (Bravery et al., 2017).

Based on the average of eight meta-analytic estimates, Schmidt et al. (2008) presented a validity coefficient for cognitive ability of .65. In practical terms, this translates to "65% of the gain in job performance that would be realized with perfectly accurate selection" (Schmidt et al., 2016, p. 14). However, the predictive validity of cognitive ability varies by job complexity. Hunter, Schmidt, and Le (2006) found that the validity coefficient was .74 for professional and managerial jobs, .66 for medium complexity jobs including skilled blue-collar jobs and mid-level white-collar jobs, and .39 for unskilled jobs. In addition to predicting job performance, cognitive ability also has high predictive validity for job training performance, with an average validity coefficient of .67 (Schmidt et al., 2008).

Predictive validity of Gf and Gc. It is important to note that much of the research concerning validity evidence of cognitive ability tests has focused on measures of *g* rather than *Gf* or *Gc* individually. Several studies have found validity evidence for other types of abilities including perceptual ability (Hunter & Hunter, 1984) verbal ability, numerical ability, spatial-mechanical ability, and memory (Salgado et al., 2003). Additionally, Postlethwaite (2011) conducted a meta-analysis to compare the validity evidence for *Gf* and *Gc*. He found that *Gc* was a more valid predictor of job performance, with a validity coefficient of .49, than *Gf*, with a coefficient of .27, and *g*, with a validity coefficient of .43. More recently, Dilchert (2017) conducted a second-order meta-analysis of the predictive validity of *g* as well as specific abilities and found validity coefficients for quantitative ability tests (i.e., measures of *Gf*) of .50 for job proficiency criteria and .39 for supervisory ratings. Validity coefficients for tests of verbal ability (i.e., measures of *Gc*) were .39 for job proficiency criteria and .51 for supervisory ratings. The validity coefficient for *g* was .53. Thus, findings across empirical studies and meta-analyses indicate that although specific abilities may be good predictors of job performance, *g* provides the strongest predictive validity evidence (Dilchert, 2017; Salgado, 2017). However, concerns about mean subgroup differences in *g* highlight the need to further investigate the validity evidence for specific abilities, as in some cases emphasizing these abilities over *g* has the potential to reduce adverse impact (Ployhart & Holtz, 2008).

Mean subgroup differences in cognitive ability test scores. Just as it is widely agreed that cognitive ability is one of the best predictors of job performance, there is also an acceptance of the fact that cognitive ability tests often result in subgroup differences between minority and non-minority applicants (Hough, Oswald, & Ployhart, 2001; Ones, Dilchert, Chockalingam, & Salgado, 2017; Outtz, 2011; Salgado, 2017; Schmidt et al., 2016). The most common method of quantifying subgroup differences is by calculating Cohen's *d*, which is a measure of a standardized difference between two means. Values greater than or equal to .80 are generally considered to be large effects, values around .50 are considered to be moderate, and those below .20 are considered to be small (Cohen, 1977). However, the practical implications of *d* values are a function of a selection system as a whole, rather than a single test. Thus, in addition to subgroup differences, the selection ratio and the method of combining and using test scores to make decisions should be taken into account (Sackett & Roth, 1996). Consequently, when small selection ratios are used (i.e., only a very small percentage of applicants are actually selected into the organization), even small *d* values below .20 can lead to adverse impact (Ones et al., 2017).

Differences based on race/ethnicity. Much of the discrimination literature has focused on race/ethnic differences, comparing Caucasian (White) candidates and employees to African

Americans (Blacks), Hispanic/Latino Americans, Asian Americans, and Native Americans/Pacific Islanders (Ones et al., 2017). Roth, Bevier, Bobko, Switzer, and Tyler (2001) conducted a large meta-analysis and found a *d* value of 1.10 between Whites and Blacks and .72 between Whites and Hispanics for *g*. However, the *d* values for specific abilities were lower for both groups, with values of .83 and .74 for verbal and numerical abilities respectively between Whites and Blacks, and values of .40 and .28 between Whites and Hispanics. Hough and colleagues (2001) provided a review of the literature on subgroup differences and reported similar *d* values of 1.00 between Whites and Blacks for *g* and average *d* values of .67 for tests of verbal ability and .69 for tests of quantitative ability. However, Bobko and Roth (2013) pointed out two major methodological issues in existing estimates of White-Black subgroup differences. Specifically, *d* is often computed using incumbent samples rather than applicant samples and researchers often confound methods and constructs (i.e., report separate values for different methods of selection that are all different measures of the same construct). In an effort to correct for these methodological issues, the researchers presented more accurate estimates of *d* values between Whites and Blacks of .72 for jobs of moderate complexity and .86 for jobs of low complexity.

Differences based on sex. Although sex differences in cognitive ability test performance have been examined to some extent in the literature, existing indicate that there are no meaningful differences between men and women (Dilchert, 2017; Ones et al., 2017; Salgado, 2017). Ployhart and Holtz (2008) reported a meta-analytic *d* value of 0.00 for male-female comparisons of cognitive ability test performance. Hough and colleagues (2001) also reported a mean *d* value for males and females of 0.00 for *g* but found small differences in tests of verbal ability that favored women (*d* = -.10) and in tests of quantitative ability that favored men (*d* = .20).

Differences based on age. Although there is a large body of literature examining age differences in cognitive ability (to be discussed in further detail in later sections of this paper), little research has focused on subgroup mean differences in cognitive ability test performance based on age (Dilchert, 2017; Fisher et al., 2017; Ones et al., 2017; Truxillo et al., 2014). Avolio and Waldman (1994) found that differences in *g* across age groups were generally small until the age of 65, but they reported stronger negative relations for tests of *Gf*. Based on these same data and using a reference age group of 20-34 years, Ones and colleagues (2017) reported *d* values of .33 for the age range of 35-44, .55 for the age range of 45-54, and .80 for the age range of 55-65 for measures of *g*. These values differed for tests of verbal ability (*d* = .26, .35, .49 for the respective age groups) and for tests of numerical ability (*d* = .36, .59, .71 for the respective age groups). Based on their summary of the literature related to subgroup differences, Hough et al. (2001) concluded that measures of *g* are likely to result in adverse impact based on age, and they reported a mean *d* value of .40 for younger (i.e., based on an average age of 30 years) compared to older (i.e., based on an average age of 50 years) workers. Finally, in a study of job applicants to executive positions, Klein, Dilchert, Ones, and Dages (2015) found that although older workers, particularly those older than 55, scored lower on tests of *g* compared to the reference group of 20-34 years of age (*d* = .32), older workers in all three age groups (i.e., 35-44, 45-54, 55-65) performed better on tests of verbal ability (*d* = -.36, -.49, -.76). However, workers in these older age groups also had lower scores on measures of *Gf* including figural reasoning (*d* = .09, .20, .36) and inductive reasoning (*d* = .51, .80, 1.03).

Differential validity and prediction. Beyond subgroup mean differences, differential validity and differential prediction have also been topics of interest in relation to cognitive ability testing. Differential validity refers to "differences in the correlation between the test and the criterion (i.e., validity) between subgroups" (Berry, 2015, Table 1). For example, if the correlation between cognitive ability test scores and job performance was .50 for applicants under the age of 40 and .40 for applicants age 40 and older, the test would show differential validity. Differential prediction, refers to "differences in regression equations between subgroups" (Berry, 2015, Table 1). For example, if the regression equation for cognitive ability test scores predicting job performance is $Y = 1.5 + .50X$ for applicants under the age of 40 and $Y = 1.0 + .40X$ for applicants age 40 and older, the test would show differential prediction based on both the different intercepts and the different slopes. Differential prediction is also referred to as test bias, which is defined in the Society for Industrial and Organizational Psychology's *Principles for the Validation and Use of Personnel Selection Procedures* (2003) as an instance in which "for a given subgroup, consistent nonzero errors of prediction are made for members of the subgroup" (p. 32). Given the focus of the current study, the examples above use age as the basis for differential validity and prediction. However, although a substantial amount of research has examined these factors based on race/ethnic differences and, to a lesser extent, sex differences, there have not been any investigations of differential validity and prediction based on age (Fisher et al., 2017; Ones et al., 2017).

Cognitive Ability and Age

Empirical evidence. Although early investigations of the relation between age and cognitive abilities were characterized by inconsistent findings (Horn & Cattell, 1967), a large body of cognitive research supports the notion that age trajectories of cognitive abilities differ for

Gf and *Gc* such that *Gc* continues to increase as individuals age whereas *Gf* declines with age (e.g., Peeters & van Emmerik, 2008; Salthouse, 2012; Schaie, 1994; Verhaeghen & Salthouse, 1997). Evidence from cross-sectional studies indicates that *Gf* peaks in one's early 20s and steadily declines thereafter, with some research findings suggesting accelerated declines after age 60 (Horn, 1975; Salthouse, 2009, 2012). In contrast, *Gc* typically increases until at least age 60 (Kanfer & Ackerman, 2004; Salthouse, 2009). It is important to note the limitations of cross-sectional research, however. For example, known as the Flynn effect, Flynn (1987) observed a phenomenon wherein cognitive ability scores increase by an average of 20 points per generation. Additionally, these generational differences are largely reflected in measures of *Gf* as opposed to *Gc* (Flynn, 1987, 1998; Raven, 2000). Thus, observed age differences in *Gf* in cross-sectional studies may be partially attributed to cohort differences.

Results of longitudinal studies of cognitive ability, such as the Seattle Longitudinal Study (Schaie, 1994), are largely consistent with results from cross-sectional studies regarding age-related declines (Horn, 1998). However, on average, the declines in *Gf* found in longitudinal studies begin at slightly older ages than those found in cross-sectional studies, perhaps indicating practice effects when participants complete the same cognitive ability measures at multiple time points (Horn & Donaldson, 1980). Additionally, in a longitudinal study of Scottish participants, Deary and colleagues (Deary, Whalley, Lemmon, Crawford & Starr, 2000; Deary, Whiteman, Starr, Whalley, & Fox, 2004) found an adjusted correlation of .73 between scores on a cognitive ability test taken at age 11 and again at age 77. Thus, within-person differences in cognitive ability across the lifespan are largely stable.

Theoretical background. Salthouse (1996) proposed a processing speed theory to explain declines in *Gf* associated with age. The theory posits that as individuals age, the speed with which they are able to execute basic mental operations declines. The results from many studies have supported the tenets of processing speed theory. For example, when controlling for processing speed, the relation between age and several cognitive tasks including free recall (Bryan & Luszcs, 1996), cued recall (Park et al., 1996), paired associates (Salthouse, 1994), and reasoning (Salthouse, Fristoe, McGuthry, & Hambrick, 1998) are substantially attenuated. Additionally, Verhaeghen and Salthouse (1997) conducted a meta-analysis of cross-sectional studies and found that age-related declines in *Gf* and processing speed shared more than 70% of the common variance. In a 4-year longitudinal study, however, Zimprich and Martin (2002) reported a shared variance of 28%. Nevertheless, the evidence supports the notion that age-related declines in *Gf* are at least partially driven by declines in processing speed (Finkel, Reynolds, McArdle, & Pedersen, 2007; Diamond, 2013; Salthouse & Madden, 2013). In addition to the effects of processing speed, Park (2000) proposed that working memory functions, inhibitory function, and sensory function are all mechanisms by which age leads to decreases in *Gf*. Further, McCabe, Roediger, McDaniel, Balota, and Hambrick (2010) reported evidence that working memory capacity and executive functioning shared a common underlying cognitive ability, referred to as executive attention, that was significantly negatively correlated with age and distinct from processing speed. Finally, Fisher et al. (2017) summarized the neurocognitive evidence supporting theoretical explanations for decreases in *Gf*, including decreases in orbitofrontal and prefrontal white matter associated with declines in *Gf* levels of healthy older adults (Raz et al., 2008).

Although much of the research regarding age-related increases in *Gc* has been empirically driven, Umanath and Marsh (2014) offered a theoretical perspective on this well-established relation. The basis of the theory is that older adults have had more opportunity to gain knowledge through their various life experiences, and this knowledge remains available in memory (Brod, Werkle-Bergner, & Shing, 2013) and typically continues to increase as they age (Cornelius & Caspi, 1987; Staudinger, Cornelius, & Baltes, 1989). The ability for older adults to retain and apply acquired knowledge has been demonstrated in a variety of situations, including those relevant to the workplace (Charness, 1981; Perlmutter, 1988; Salthouse, 1994; Waldman & Avolio, 1986, 1993). The research findings related to the relation between age and job performance, summarized in the next section, provide further evidence to support this theory.

Age and Job Performance

Empirical evidence. Although research shows mixed results regarding the relation between age and job performance, the general conclusion is that age-related differences range from negligible to small (Salthouse, 2012). Job-specific exceptions do exist; for example, the U.S. has imposed a mandatory retirement age of 56 for air traffic controllers due to the overwhelming evidence suggesting age-relate declines in performance (Cobb, Lay, & Bourdet, 1971; Heil, 1999; Salthouse, 2012; Trites & Cobb, 1963, 1964). Several meta-analyses have investigated the relation between age and job performance across job types. The first of these was conducted by Waldman and Avolio (1986) who found that age was positively related to job performance when the criterion was measured by productivity (.27), but negatively related to supervisor ratings of performance (-.14). In contrast, McEvoy and Cascio (1989) reported a meta-analytic estimate of .06 for the relationship between age and job performance and did not find moderating effects for performance rating type. Sturman (2003) reported an estimate of .03 and found support for an inverted-U-shape for the relationship, wherein age was positively related to performance for those under 49 years of age but negatively related to performance for those 49 years and older.

The conflicting results regarding the relation between age and job performance can be partially explained by differences in sample characteristics (e.g., age dispersion) and data collection methods (e.g., cross-sectional vs. longitudinal design; Ng & Feldman, 2008). However, it is also important account for both the type of cognitive abilities measured and the requirements of different jobs. Given the evidence discussed in previous sections regarding the negative relation between age and *Gf*, age should be negatively related to job performance for jobs that require fluid abilities such as inductive or quantitative reasoning (Avolio & Waldman, 1994). Conversely, the age-related increases in *Gc* associated with acquired job experience and knowledge should give older workers an advantage regarding jobs involving deductive reasoning or professional experience (Masunaga & Horn, 2001). Indeed, several theories have been proposed to explain the mechanisms by which older workers may compensate for declines in *Gf* with more reliance on *Gc*.

Theoretical background. One theory that provides a useful framework for understanding the differential findings regarding age and job performance based on different types of cognitive abilities is selection, optimization, compensation (SOC) theory (Baltes & Baltes, 1990). Originally proposed as a model of successful aging, the SOC model suggests three strategies that older adults can use to manage limitations in personal resources. Selection involves setting goals according to the importance that an individual places on different domains. Optimization refers to activities aimed to enhance or maintain the strategies used to achieve the selected goals. Finally, compensation involves changing or adapting these goals or strategies when necessary to maintain a certain level of functioning. Thus, older workers may compensate for declines in certain cognitive abilities by shifting jobs (Warr & Pennington, 1994), reducing the range of domains in which they choose to maintain expertise (Salthouse, 2012), or delegating certain cognitively loaded tasks to others (Birren, 1969). Baltes and Baltes (1990) ultimately proposed that older adults could make up for declines in cognitive mechanics (i.e., fluid abilities) by capitalizing on increased task-related knowledge. Salthouse (1984) provided further evidence to support this theory. In a study of older and younger typists, he found that older typists had a slower typing speed than younger typists; however, the older typists still outperformed the younger typists because they had learned to read further ahead in the text that was to be typed. Thus, the older typists had developed a strategy to compensate for their age-related loss in tapping speed (Baltes & Baltes, 1990). Additionally, Castel (2007) summarized literature suggesting that older adults are more likely to rely on high-value information along with prior knowledge to compensate for age-related declines in memory. That is, as individuals age, they are able to strategically focus on information that they deem important to remember.

Remaining Gaps in the Literature

Although the topics of age, cognitive ability, and job performance have been studied extensively, several unanswered questions remain, especially regarding the intersection of these three research areas. These gaps in the literature were highlighted by Fisher and colleagues (2017) and Truxillo and colleagues (2014). First, differential prediction and differential validity have not been studied in the context of age and cognitive ability testing for selection purposes. Second, although adverse impact based on age has been investigated in a handful of studies (e.g., Klein et al., 2015), they called for further examination of this issue. Beyond examining these differences between employees and candidates under 40 years of age and over 40 years of age, they also pointed to the value in investigating age discrimination using age as a continuous variable. Third, Morgeson, Reider, Campion, and Bull (2008) specifically called for more research on age discrimination using field samples rather than student or convenience samples.

Present Studies

The goal of the two present studies is to further investigate age discrimination in relation to cognitive ability testing for selection purposes. I draw from theories across the aging, cognitive ability, and job performance literatures to specifically address the gaps in the current literature discussed in the previous section. In both studies, I used field samples of job incumbents across a range of ages to examine age differences in cognitive ability test performance and differential validity and prediction based on age. In the first study, fluid cognitive abilities (i.e., numerical reasoning, figural reasoning, and deductive reasoning) were measured via a timed paper-and-pencil test. In the second study, crystallized abilities were measured via an untimed online assessment of job-related knowledge.

Despite the differences in the measures, there are several important similarities between the studies that should be considered, especially when comparing the results. Both samples were drawn from the same organization, although the data were collected several years apart. All participants across both studies were incumbents in management positions at similar job levels and were selected to participate in concurrent validation studies for new selection tests. Finally, although different measures of job performance were used, both studies relied on supervisory ratings of job performance as the criterion measure.

Study 1

Hypotheses

Given previous empirical evidence of age-related declines in fluid cognitive abilities, I expect that older test-takers will perform worse than younger test-takers on a test of three types of fluid abilities—numerical reasoning, deductive reasoning, and figural reasoning. Additionally, processing speed theory (Salthouse, 1996) suggests that tests with specified time limits will pose constraints for older test-takers who may work through problems at a slower rate than younger test-takers. It is important to consider this from a legal perspective, as I expect this will lead to age-based adverse impact, especially when overall selection ratios are low, such that test-takers over the age of 40 will be selected at a rate that is less than 80% of the rate for those under 40 years of age. However, I also expect to find a negative, linear relation between age and cognitive ability test performance.

Hypothesis 1: *The tests of a) numerical reasoning, b) figural reasoning, and c) deductive reasoning will result in adverse impact based on age.*

Hypothesis 2: *There will be a negative, linear relation between age and test performance on a) numerical reasoning, b) figural reasoning, and c) deductive reasoning.*

Detection of adverse impact depends on a number of factors including sample size, overall selection rates, and the method used for adverse impact analysis. Thus, each of these factors were taken into account and discussed in the analyses and results sections below.

Theories of job performance and cognitive ability suggest that fluid cognitive abilities predict job performance because those with higher levels of *Gf* are more likely to acquire more job knowledge at a faster rate when they are hired. However, the empirical evidence of age-related increases in job knowledge suggests that older adults who have experienced declines in

Gf bring previously acquired job knowledge with them once hired. Thus, tests of *Gf* should be more likely to predict performance for younger candidates because it is a measure of their potential to acquire knowledge, whereas older candidates who do not perform as well on these tests due to declines in *Gf* already possess higher levels of acquired knowledge that will positively influence their job performance. I expect that the test of fluid abilities will result in both differential validity, such that test performance will be more strongly related to job performance for younger test-takers, and differential prediction, such that it will underpredict job performance for older test-takers.

Hypothesis 3: *The tests of fluid cognitive abilities will show differential validity based on age such that the validity coefficients will be higher for younger adults (i.e., under 40 years of age) compared to older adults (i.e., age 40 years and older). This will be the case for the a) numerical reasoning, b) figural reasoning, and c) deductive reasoning sections of the test.*

Hypothesis 4: *The tests of fluid cognitive abilities will show differential prediction based on age such that the intercept in the regression equation for test performance predicting job performance will be lower for younger adults (i.e., under 40 years of age) compared to older adults (i.e., age 40 years and older). This will be the case for the a) numerical reasoning, b) figural reasoning, and c) deductive reasoning sections of the test.*

Method

Participants. Participants were 214 incumbents of various mid-level management roles at a large consumer goods company in the United States. The sample consisted of 91 females (42.5%) and 123 males (57.5%) ranging in age from 25 to 70 with a median age of 37 (M = 38, SD = 7.98). 63% of participants (N = 135) were under the age of 40 and 37% (N = 79) were 40 years of age or older. Within the sample, 56% were White/Caucasian, 12% Black/African American, 14% Hispanic, and 15% Asian, with 2% of the sample not providing race/ethnicity data. They also represented a variety of departments within the organization including manufacturing (26%), research and development (21%), finance and accounting (17%), information technology (13%), and marketing (11%).

Procedure. All participants were randomly selected to participate in a concurrent validation study for a cognitive ability test to be used for selection purposes by the organization. Participation in the validation study was voluntary and participants were told that their data would not be used for administrative purposes or be shared on an individual level with anyone in the organization. Employees who chose to participate were asked to take a paper-and-pencil cognitive ability test consisting of 40 items broken into three sections—numerical reasoning, deductive reasoning, and figural reasoning. The time limit for the test was 65 minutes and all participants took the test in a proctored setting. At the same time that the test data were collected, the managers of all invited incumbents were asked to complete an online survey consisting of 7 performance dimensions to indicate their employees' level of job performance. Before providing performance ratings, managers participated in a brief rater error training session during which they were educated about common rater errors including leniency (i.e., rating too positively), severity (i.e., rating too negatively), and halo error (i.e., rating according to an overall impression rather than specific to distinct performance competencies).

Measures. The following measures were used in the study:

Numerical reasoning. The numerical reasoning section of the test originally consisted of 15 multiple-choice word problems with five answer options ($\alpha = .75$). The numerical reasoning score was indicated by the total number of items that were answered correctly.

Deductive reasoning. The deductive reasoning section of the test consisted of 10 items that each included a paragraph of information and four response options (α = .66). After reading the information in each paragraph, test-takers were instructed to select the answer option that represented either the only valid statement that could be logically concluded from the information provided in the paragraph or the only invalid statement that could not be logically concluded from the information provided in the paragraph. The deductive reasoning score was indicated by the total number of items in the section that were answered correctly.

Figural reasoning. The figural reasoning section consisted of 15 items that each included a series of figures with varying properties (e.g., shapes, patterns, shades, etc.; α = .78). Test-takers were asked to select the figure that best completed the series out of five possible options. The figural reasoning score was indicated by the total number of items in the section that were answered correctly.

Job performance. Job performance was measured via supervisory ratings on six dimensions of performance—gathering information, innovation, reviewing and analyzing information, strategic and operational agility, decision making, and potential. For each dimension, managers were presented with behaviorally anchored rating scales and asked to provide a rating on a scale of 1 (needs a lot of improvement) to 7 (demonstrates real strength). The original supervisory measure consisted of seven dimensions, but the results of a confirmatory factor analysis indicated that removing the seventh dimension, adaptability, significantly improved the fit of a unidimensional model (CFI = .94; TLI = .90; RMSEA = .09). Thus, a composite score consisting of an average of the six other performance ratings was used for the analyses (α = .78).

Statistical Analyses

Adverse impact analysis. Although the four-fifths rule was provided as a guideline for determining adverse impact (EEOC, 1978), it was not intended to be a strict legal guideline given that the impact ratio (i.e., selection rate for the nonminority group divided by the selection rate for the protected group) can be heavily influenced by sample size (Zedeck, 2010). That is, subgroup differences that do not violate the four-fifths rule can still be ruled as indicative of adverse impact if the differences are statistically significant and differences that violate the four-fifths rule may not constitute adverse impact if they are based on small sample sizes or are not statistically significant. Jacobs, Murphy, and Silva (2013) conducted a simulation study in which the sample size, overall selection ratio, and group difference (i.e., *d* value) were varied and discovered that when statistical significance tests are used to determine adverse impact, sample size was more likely to impact this determination than the size of the difference in hiring outcomes for subgroups. The researchers called for regulatory agencies to take this into consideration regarding the process for determining adverse impact; however, the four-fifths rule and statistical significance testing are still the most commonly used practices for determining adverse impact in court (Oswald, Dunleavy, & Shaw, 2017).

I conducted several different analyses to determine whether adverse impact on the basis of age would be likely to occur for the test of *Gf*. These analyses were based on the procedures often used in compliance proceedings (Roth, Bobko, & Switzer, 2006). Additionally, I conducted each of these analyses using a moderate selection scenario with an overall selection rate of 50% and an extreme selection scenario with an overall selection rate of 10% (as used in Oswald et al., 2017). The conclusions based on these analyses are also based on the assumptions that the test is used in a single-hurdle context (i.e., candidates are selected solely based on the test scores) and a top-down selection approach (i.e., those with the highest test scores are selected).

Four-fifths (80%) rule. I determined the presence of adverse impact based on the four-fifths rule by dividing the selection ratio for participants 40 years of age and older by the selection ratio for participants under the age of 40. Impact ratios that were less than .80 indicated the presence of adverse impact based on the four-fifths rule, and values greater than or equal to .80 indicated the absence of adverse impact. I calculated impact ratios based on both the moderate and extreme selection scenarios (i.e., 50% and 10% overall selection ratios).

Tests of statistical significance. In the context of legal proceedings, when analyses based on the four-fifths rule indicate the presence of adverse impact, the next step is typically to conduct statistical significance testing (Roth et al., 2006). Switzer, Roth, and Rosopa (2016) examined a variety of statistical significance testing techniques under various conditions and found that Fisher's exact test with Lancaster's Mid-P correction (Biddle & Morris, 2011) and the Z-test for proportions were the two methods most likely to perform adequately across scenarios. Additionally, Collins and Morris, 2008 found that the Z-test for proportions was particularly appropriate when using small sample sizes. Thus, I computed values based on both Fisher's exact test as well as the Z-test. Statistical significance was indicated by p values less than .05.

For each test of statistical significance, I also conducted a sensitivity analysis to determine the largest effect size that could be detected given the sample sizes of each subgroup, the alpha level, and statistical power. The alpha level represents the probability of rejecting the null hypothesis when the null hypothesis is true, and power represents the probability that the null hypothesis will be rejected when the null hypothesis is false (Cohen, 1988).

Indicators of practical significance. Oswald et al. (2017) offered several practical significance measures that can be used in conjunction with the four-fifths rule and statistical significance testing, although the researchers acknowledged that these have not been widely used in legal proceedings. First, the phi coefficient is calculated based on a 2 x 2 frequency matrix containing subgroups (i.e., under 40 years versus 40 and older) and outcome (i.e., selected based on test score versus not selected based on test score). Phi represents the square root of the chi-square value divided by the total sample size. The squared phi coefficient represents the variance accounted for in the outcome by subgroup membership. Based on rule of thumb guidance from Murphy and Jacobs (2012), a phi coefficient of .10 or higher (i.e., indicating that subgroup membership accounts for at least 1% of the variance in selection outcomes) indicated practically meaningful adverse impact. I also calculated 95% confidence intervals to indicate the precision of the estimate (Bonnet & Price, 2007).

The odds ratio is another suggested practical significance measure that also relies on the 2 x 2 frequency matrix. Odds ratios represent the odds of an individual from one subgroup being selected versus the odds of an individual from another subgroup being selected. Gastwirth (1988) suggested that odds ratios greater than or equal to 1.4 or less than or equal to .71 indicate meaningful differences between selection ratios for the two subgroups. I calculated the odds ratios based on the selection ratios for those under 40 years of age and those 40 years of age and older as well as confidence intervals for the odds ratios as recommended by Oswald and colleagues (2017).

A final practical significance measure that I calculated is the *h* statistic (Cohen, 1988), which represents a standardized difference in selection ratios. When interpreting the *h* statistic, the same guidelines as the *d* value were be used, wherein *h* values less than .20 were considered small, those around .50 were considered medium, and those greater than .80 were considered large (Cohen, 1988).

Linear relations between age and cognitive ability. To examine linear relations between age and cognitive ability, I conducted a multiple linear regression analysis. First, I regressed the cognitive ability test score on race/ethnicity and gender, as these demographic variables are associated with subgroup differences in cognitive ability test scores. Next, I added age to the regression model and examined the change in R^2 to assess whether age accounted for a significant proportion of the variance in test scores above and beyond that accounted for by race/ethnicity and gender.

Differential validity. I examined differential validity by calculating the test-criterion correlation coefficients for each subgroup and using the Fisher r-to-Z transformation to calculate the statistical significance of the differences in correlations between subgroups. I also calculated Cohen's *q*, which is a measure of the effect size of differences in correlation coefficients.

Corrections for unreliability and range restriction. Given that participants in the sample were incumbents who had already been selected into the organization based on previous assessments, it is likely that range restriction occurred such that the sample correlation was not representative of the population correlation between the predictor and the criterion. This is a common issue in validity analyses that can lead to attenuated validity coefficients (Alexander, Alliger, & Hanges, 1984; Schmidt & Hunter, 1998; Sackett & Yang, 2000). Sackett and Yang (2000) presented several different methods of correcting for range restriction depending on the information available. Many of these methods require the researcher to know the unrestricted population variance for the predictor (i.e., test), criterion (i.e., performance), or both. I followed the method originally proposed by Alexander and colleagues (1984) that can be used when the unrestricted variance is unknown for both variables. Following this method, I calculated Cohen's (1959) ratio based on the variance, mean of the test variable in the sample, and the lowest observed score of the test variable in each sample. Next, I referred to the table provided by Alexander et al. (1984) to obtain the z-score associated with the point of truncation as well as the estimate of the standard deviation in a normal distribution truncated at that point. I then used these values in Thorndike's (1949) Case 2 formula, which relies on the restricted standard deviation of the predictor as well as the unrestricted standard deviation (as estimated using Alexander et al.'s (1984) method), and the observed correlation to produce an estimate of the corrected correlation.

Similarly, unreliability in the predictor and criterion measures also leads to attenuation of the correlation coefficient (Ghiselli, 1964; Lord & Novick, 1968). The disattenuated correlation is calculated by dividing the correlation between the predictor and the criterion by the square root of the product of the reliabilities of each measure (Murphy & Davidshofer, 1988).

I corrected the validity coefficients for each subgroup and for each section of the test using the procedures above. Stauffer and Mendoza (2001) posited that when selection is made on the basis of the restricted predictor (i.e., the observed test score), correlation coefficients should be corrected for unreliability after correcting for range restriction. Thus, I first corrected for range restriction and then corrected for unreliability in both the predictor and the criterion.

Differential prediction. The typical approach for examining differential prediction is to conduct moderated multiple regression by regressing job performance first on test scores, then on the subgroup variable, and finally on the test-subgroup interaction term and to examine the significance of the interaction term (Cleary, 1968; Drasgow & Kang, 1984). However, Berry (2015) suggested the simpler procedure of generating regression equations separately for each subgroup and comparing the slopes and intercepts. Thus, I conducted differential prediction analyses using both approaches for each test. Based on Berry's (2015) method, I generated the regression equation for job performance regressed on cognitive ability test score for the portion of the sample that was younger than 40 years of age and generated this same regression equation for the portion of the sample that was 40 years of age or older. I then compared the regression slopes and intercepts for the two groups. Differential validity was indicated by different regression slopes between subgroups, and differential prediction was indicated by either different regression slopes or different intercepts between subgroups. I then examined the age subgroup-test score interaction terms using the moderated multiple regression method to evaluate the statistical significance of these differences.

Study 1: Results

Descriptive Statistics

Descriptive statistics and correlations for all study variables are reported in Tables 1 and 2. I also examined the frequency distribution of each of the variables of interest to assess normality and identify any outliers. Overall, the distribution for age was slightly right-skewed such that a larger proportion of the sample was younger. In addition, the distribution for performance was left-skewed such that a larger proportion of the sample received higher performance ratings. As well, the distributions of scores for all sections of the test were left-skewed such that a larger proportion of the sample achieved higher scores on the tests, with the largest skew for the numerical reasoning section. I conducted cubic and square transformations of the numerical reasoning score data to attempt to mitigate the effects of the skewness; however, the transformed data did not produce different results for the statistical analyses described below.

Thus, all results are based on the original data rather than the transformed data. No outliers were identified for any of the variables based on Grubbs' test for outliers (Grubbs, 1969).

Adverse Impact Analysis

Numerical reasoning. I found moderate differences in test scores for the numerical reasoning between the subgroup below the age of 40 and those age 40 years and older, such that the older group attained lower scores on average (*d* = .45). Given that 21% of the total sample achieved the highest possible score on the numerical reasoning section, the extreme selection scenario was based on a 21% overall selection rate. Adverse impact based on age was indicated in this scenario, with an adverse impact ratio of .79, indicating that those 40 years and older would have been selected at a rate that was 79% of the rate at which candidates under 40 years of age would have been selected. The differences in selection ratios were not statistically significant based on Fisher's Exact test ($p = .49$) and the Z test of proportions ($Z = -0.81$). Additionally, the results of a sensitivity analysis indicated that given the selection ratio of .18 for the older subgroup, the smallest possible selection ratio for the younger subgroup that would have allowed me to detect a statistically significant difference between proportions was .45. The actual selection ratio of .22 for the younger subgroup did not meet this criterion. Additionally, tests of practical significance did not meet sufficient criteria to indicate adverse impact based on the phi value ($\varphi = .05$), the odds ratio ($OR = 1.33$), or the *h* statistic ($h = .11$; Table 3).

Deductive reasoning. There were moderate differences in test scores for the deductive reasoning section between the subgroup below the age of 40 and those age 40 years and older, such that the older group attained lower scores on average (*d* = .45). At an extreme overall selection rate of 12%, adverse impact was indicated with an adverse impact ratio of .51, but differences in selection ratios between groups were not statistically significant based on Fisher's

Exact test (p = .13) or the Z test of proportions (Z = -1.75). The results of a sensitivity analysis indicated that given the selection ratio of .08 for the older subgroup, the selection ratio of the younger subgroup (i.e., .15) did not meet the minimum that would have allowed me to detect a statistically significant difference between proportions (i.e., .28). However, a phi coefficient of .11 indicated that 1.1% of the variance in the selection outcome was accounted for by subgroup membership, meeting the minimum criterion for practical significance as suggested by Murphy and Jacobs (2012). An odds ratio of 2.12 also indicated practical differences in selection ratios; however, the 95% confidence interval (.81 – 5.52) included the suggested minimum indicator of practical significance (i.e., 1.4). Finally, an h statistic value of .23 suggested that small standardized differences between selection ratios were present (Table 3).

Figural reasoning. There were moderate differences in test scores for the figural reasoning section between the subgroup below the age of 40 and those age 40 years and older, such that the older group attained lower scores on average (d = .46). Adverse impact was indicated at an overall selection ratio of 9%, with an adverse impact ratio of .61. However, the differences in subgroup selection ratios were neither statistically significant (Fisher's Exact p = .46; Z = -1.09) nor practically significant (φ = .07; OR = 1.7, 95% CI = .59-4.95; h = .15; Table 3). Again, the results of a sensitivity analysis indicated that given the selection ratio of .08 for the older subgroup, the selection ratio of the younger subgroup (i.e., .10) did not meet the minimum that would have allowed me to detect a statistically significant difference between proportions (i.e., .25). .

Exploratory adverse impact analyses. Although the legal cutoff for age as a protected class is 40 years, I conducted the same adverse impact analyses after increasing the age cutoff to 45 years. The purpose of these additional analyses was to explore the implications of adjusting the legal standard as well as to gain more insight as to the practical implications of tests that may discriminate against older workers and prevent organizations from hiring experienced candidates and achieving age diversity in the workforce. It is important to acknowledge the limitations of these exploratory analyses, as adjusting the age cutoff further reduced the sample size for the older subgroup, further diminishing statistical power.

Numerical reasoning. Adverse impact was indicated under the 21% and 55% selection ratios with adverse impact ratios of .42 and .57 respectively. Given a 21% selection ratio, the difference in subgroup selection ratios was not statistically significant based on the Fisher's Exact test (p = .08) but was significant based on the Z test of proportions (Z = -2.53). The phi coefficient and h statistic both met the threshold for practical significance (φ = .13; h = .37); however, the 95% confidence interval for the odds ratio did not indicate meaningful differences (OR = 2.78, 95% CI = .93-8.30; Table 4).

Deductive reasoning. Adverse impact was not indicated under any of the three selection scenarios (SR = 12%, 53%, 91%). In fact, the selection ratios for each age subgroup were almost exactly the same for the numerical reasoning section (Table 4).

Figural reasoning. Adverse impact was indicated under the 9% and 44% selection scenarios with adverse impact ratios of 0.0 and .74 respectively. Tests of statistical significance could not be conducted and practical indicators could not be calculated given the 0% selection ratio for the older subgroup. Given a 44% selection ratio, the difference in subgroup selection ratios was not statistically significant based on the Fisher's Exact test (p = .22) nor the Z test of proportions (Z = -1.55). The phi coefficient and h statistic suggested that the selection ratios were meaningfully different (φ = .10; h = .25); however, the 95% confidence interval for the odds ratio did not meet the minimum suggested threshold (OR = 1.67; 95% CI = 0.81-3.38; Table 4).

Linear Relations between Age and Cognitive Ability

I conducted three separate multiple linear regression analyses to examine linear relations between age and each measure of cognitive ability, first regressing the score for each section of the test on the control variables (i.e., race/ethnicity and gender) and then adding the predictor of interest (i.e., age at the time the test was taken). There were small, negative relations between age and scores on the numerical reasoning (β = -.28, p < .01), deductive reasoning (β = -.24, p < .01), and figural reasoning (β = -.29, p < .01) sections of the test when controlling for race/ethnicity and gender. Additionally, for each section of the test, adding age to the regression model significantly increased the R^2 values compared to the original models with just the control variables (Table 5). The results of a sensitivity analysis showed that the observed effect sizes (f^2 = .10, .42, .36) exceeded the smallest effect that could be detected given the sample size, alpha level, and statistical power (f^2 = .06).

Differential Validity

I examined differential validity by calculating the test-criterion correlation coefficients for each subgroup. The correlations between the test score and performance measure were r = .09 and r = .17 for the younger and older subgroups respectively for the numerical reasoning section, r = .15 and r = .12 for the deductive reasoning section, and r = -.03 and r = .16 for the figural reasoning section. Using the Fisher r-to-Z transformation, I calculated the statistical significance of the differences in correlations and found that the differences were nonsignificant for all three sections (p = .29, .42, .09). I also calculated Cohen's q to interpret the effect size of the difference and found no meaningful effects for the numerical and deductive reasoning sections but a small effect (q = .19) for the figural reasoning section in the opposite direction as hypothesized, as the validity coefficient was larger for the older subgroup than for the younger subgroup. However the results of a sensitivity analysis indicated that the smallest effect that could be detected given the sample sizes, alpha level, and statistical power was .51.

I also conducted the corrections for first range restriction and then unreliability. The corrected coefficients for range restriction were the same as the uncorrected coefficients; however, the corrected coefficients for unreliability were stronger as expected for the numerical reasoning (r = .13 for younger; r = .21 for older), deductive reasoning (r = .22 for younger; r = .16 for older), and figural reasoning sections (r = -.04 for younger; r = .20 for older).

Differential Prediction

Comparisons of the regression equations for each age-based subgroup and for each section of the test are presented in Figures 1-3. There were small differences in both the slope (Δ = .02) and intercept (Δ = .10) for the numerical reasoning section, indicating that differential validity and prediction were not present. Larger differences existed for the deductive reasoning section (Δ slope = .25; Δ intercept = 1.4) such that the slope was larger and the intercept was smaller for the subgroup under the age of 40. Finally, there were differences in the slope and intercept for the figural reasoning section in the opposite direction as hypothesized (Δ slope = .23; Δ intercept = 2.8), such that the slope was larger and the intercept smaller for the subgroup aged 40 years and older.

I also conducted three separate multiple linear regression analyses in which the performance outcome was regressed first on the test score, then on the age subgroup variable, and finally on the interaction between test score and subgroup for each section of the test. Results indicated that the interaction terms for subgroup and test score were not statistically significant for the numerical reasoning (p = .85), deductive reasoning (p = .71), or figural reasoning (p = .19) sections of the test (Table 6).

Study 1: Discussion

The results of Study 1 generally indicated that age-related differences in fluid cognitive abilities were present, as participants under the age of 40 scored between .45 and .46 standard deviation units higher than those age 40 years and older on the three measures of fluid abilities. Results of the adverse impact analyses provided mixed support for Hypothesis 1. The adverse impact ratios based on age were below the EEOC's guideline of .80 for all three fluid ability measures when the overall selection ratio was low (i.e., between 9% and 21%). However, differences in subgroup selection ratios under those conditions were not statistically significant nor practically significant based on the included practical indicators. However, an exception to this was found for the deductive reasoning section; practical indicators suggested that small differences were present. It is important to note that small sample sizes led to low statistical power for the tests of statistical significance. The sample size has meaningful practical implications, given Jacobs and colleagues' (2013) argument that sample size is more likely to impact the outcome of statistical significance tests than the actual difference in selection ratios between subgroups. Thus, given that adverse impact ratios and statistical significance testing are the most commonly used practices for determining adverse impact in court cases (Oswald et al., 2017), the same analyses with larger sample sizes would have likely indicated age-based adverse impact for these measures of fluid abilities.

Hypothesis 2 regarding the linear relations between age and the three measures of fluid abilities was supported, as small, negative relations were present when controlling for race/ethnicity and gender. This is one of the few studies that has examined the continuous relation between age and tests of fluid cognitive abilities within a selection context. The results are consistent with previous literature concerning the decline of fluid cognitive abilities over the lifespan (i.e., Schaie, 1994).

Hypothesis 3 regarding the differential validity of the fluid ability measures for different age subgroups was not supported, as the differences between the test-criterion correlation coefficients were not significant. Further, although the effect sizes for the numerical reasoning and deductive reasoning sections of the test indicated that the differences in validity coefficients were not meaningful, the effect size for the figural reasoning section indicated that there were small but meaningful differences such that the validity coefficient was larger for the older subgroup than for the younger subgroup. Additionally, although small differences in the slopes and intercepts of the test-criterion regression equations for the two age subgroups were present, the nonsignificant subgroup-test score interaction terms suggested that the differences were not meaningful. Thus, although Hypothesis 4 stated that the tests of fluid abilities would differentially predict performance for older and younger participants, the results indicated that this was not the case.

Overall, the results of Study 1 indicate that although the use of tests of fluid cognitive abilities including numerical reasoning, deductive reasoning, and figural reasoning for employee selection have the potential to lead to age-based adverse impact, especially given large sample sizes and small overall selection ratios, these tests are likely to predict performance equally for older and younger applicants. Deductive reasoning appeared to be most susceptible to age-based adverse impact of the three types of tests, as the adverse impact ratio was the smallest and two of the three practical indicators suggested that meaningful differences in selection ratios between subgroups were present. Additionally, the figural reasoning test was associated with the largest differences in the predictive validity between subgroups such that there was a stronger test-criterion relation for the older subgroup compared to the younger subgroup.

Although the existing literature examining subgroup mean differences in cognitive ability test performance based on age is limited, the current results can be compared to previous results in terms of the standardized differences between subgroups (i.e., *d* values). The difference in numerical reasoning test performance in this study (d = .45) was similar to the differences reported by Ones et al. (2017), which ranged from .36 to .71 for various age groups compared to the reference group of 20-34 years. However, there was a larger difference in figural reasoning test performance (d = .46) compared to that reported by Klein et al. (2015) who also examined differences in subgroups based on those under 40 years of age and those 40 years and older (d = .18). Although I did not find any studies that explicitly examined subgroup differences in deductive reasoning based on age, the differences found in this study (d = .45) were almost identical to those found in Klein et al.'s (2015) study for inductive reasoning (d = .46).

Study 2

Hypotheses

The focus of Study 2 is a selection measure of cognitive ability that was designed to assess domain-specific knowledge. Theoretical and empirical evidence suggests that older candidates were more likely to perform well on the test than younger candidates for several reasons. First, unlike the cognitive ability measure in Study 1, the test in Study 2 did not require test-takers to use inductive reasoning or other processes that rely on fluid abilities to solve novel problems. In fact, the test required candidates to draw from previous domain-specific work experience and knowledge to solve problems that they could be expected to encounter on the job. Thus, although older candidates likely had lower levels of fluid abilities compared to younger candidates, their previous experience and crystallized abilities should confer an advantage when taking the test. Finally, whereas the test in Study 1 posed a time limit on test-takers, the test in Study 2 did not have any time constraints. Consequently, in line with processing speed theory (Salthouse, 1996), the test in Study 2 should have eliminated any negative age-related effects associated with time constraints. Thus, I expect that age-based adverse impact will not be present for the cognitive ability test. I also expect to find a positive, linear relation between age and cognitive ability test performance.

Hypothesis 1: *The untimed test of crystallized abilities will not result in adverse impact based on age.*

Hypothesis 2: *There will be a small, positive linear relation between age and test performance on crystallized abilities.*

Whereas the test of fluid abilities in Study 1 was designed for candidates with little or no relevant job experience and contained items requiring test-takers to solve novel problems, the test in Study 2 was developed with the purpose of assessing experienced candidates and contained items requiring test-takers to use this experience to solve job-relevant problems. Thus, regardless of a test-taker's age, I expect test scores to be related to job performance and not lead to differential validity or differential prediction.

Hypothesis 3: *The test of crystallized ability will not result in differential validity based on age.*

Hypothesis 4: *The tests of crystallized ability will not result in differential prediction based on age.*

Method

Participants. Participants were 232 incumbents of mid-level management roles in the Sales department at a large consumer goods company in the United States. The sample consisted of 115 females (50%) and 117 males (50%), ranging in age from 24 to 64 years with a median age of 41 years (*M* = 42, *SD* = 11.65). Within the sample, 69% were White/Caucasian, 9% Black/African American, 10% Hispanic, and 9% Asian. Race/ethnicity data for the remaining 3% of participants were missing.

Procedure. All participants were selected to participate in a concurrent validation study for a cognitive ability test to be used for selection purposes by the organization. A representative sampling method was used such that those invited to participate in the validation study represented the larger population in terms of age, gender, and race/ethnicity. Participation in the validation study was voluntary, and participants were told that their data would not be used for administrative purposes or be shared on an individual level with anyone in the organization. A total of 341 employees received a link to the online test via email and were asked to complete it within two weeks, and 232 participants (68%) completed it within the given timeframe. There was no time limit for the test, but participants were told that it would take approximately 75 minutes to complete the test. The managers of all invited participants were asked to complete a 10-minute online survey to indicate their employees' level of job performance on several dimensions.

Measures. The following measures were used in the study:

Crystallized ability. Crystallized ability was measured by a domain-specific problem-solving exercise within the test consisting of 13 items ($\alpha = .74$). Participants were given four sample sales reports that were designed to be representative of actual reports that are typically used in the role. Each item in the exercise required participants to refer to at least one of the provided reports to draw conclusions about sales trends. Four response options were provided for each item.

Job performance. Overall job performance was measured via supervisory ratings on five dimensions of performance—leadership (5 items), innovation (3 items), productivity (2 items), execution (5 items), teamwork (5 items). Additionally, domain-specific performance was assessed based on two sales-specific dimensions—aligning with customers (3 items) and influencing customers (3 items). For each dimension, managers were presented with behavioral items that represented effective performance on the given dimension and asked to provide a rating for each behavior to describe the particular employee relative to other employees on a scale of 1 (among the worst) to 7 (among the best). A composite score for overall performance was determined by first calculating the average across the items for each dimension and then determining the average of the dimension composites to arrive at an overall performance composite score ($\alpha = .89$). Results of a confirmatory factor analysis indicated that a unidimensional model was a good fit for the dimension-level data (CFI = .99; TLI = .98;

RMSEA = .09). The same approach was taken to calculate a composite score consisting of an average across the two sales performance dimensions ($\alpha = .85$).

Statistical Analyses

Statistical analyses were conducted in the same manner as Study 1. A brief summary of these analyses follows.

Adverse impact analysis. I conducted several different adverse impact analyses for the crystallized ability test using a moderate selection scenario with an overall selection rate of 50% and an extreme selection scenario with an overall selection rate of 10%. The conclusions based on these analyses were also based on the assumptions that the test is used in a single-hurdle context (i.e., candidates are selected solely based on the test scores) and a top-down selection approach (i.e., those with the highest test scores are selected).

Four-fifths (80%) rule. I determined the presence of adverse impact based on the four-fifths rule by dividing the selection ratio for participants 40 years of age and older by the selection ratio for participants under the age of 40 based on both tests. Impact ratios that were less than .80 indicated the presence of adverse impact based on the four-fifths rule, and values greater than or equal to .80, indicated the absence of adverse impact. I calculated impact ratios based on both the moderate and extreme selection scenarios (i.e., 50% and 10% overall selection ratios).

Tests of statistical significance. I assessed statistical significance of any identified adverse impact based on both Fisher's exact test as well as the Z-test. Statistical significance was indicated by p values less than .05. I conducted sensitivity analyses to determine the smallest effect that could be detected given the sample sizes, alpha level, and statistical power.

Indicators of practical significance. I evaluated the phi coefficient, odds ratios, and the *h* statistic as indicators of practical significance of age-based adverse impact. Phi coefficients of .10 or higher, odds ratios greater than or equal to 1.4 or less than or equal to .71 indicated meaningful differences between selection ratios for the two subgroups. Additionally, *h* values less than .20 were considered small, those around .50 were considered medium, and those greater than .80 were considered large.

Linear relations between age and test performance. To examine linear relations between age and cognitive ability, I conducted a multiple linear regression analysis. First, I regressed the cognitive ability test score on race/ethnicity and gender, as these demographic variables are associated with subgroup differences in cognitive ability test scores. Next, I added age to the regression model and examined the change in R^2 to assess whether age accounted for a significant proportion of the variance in test scores above and beyond that accounted for by race/ethnicity and gender.

Differential validity. I calculated the correlation between test scores and the performance measure for each age-based subgroup and calculated the statistical significance of the differences using Fisher's r-to-Z transformation as well as the effect size of the difference using Cohen's *q*. I also calculated the corrected coefficients for unreliability and range restriction.

Differential prediction. For the portion of the sample that was younger than 40 years of age, I generated the regression equation for job performance regressed on crystallized ability test score. I then generated these same regression equations for the portion of the sample that was 40 years of age or older. I compared the regression slopes and intercepts for the two groups. Differential validity was indicated by different regression slopes between subgroups and differential prediction would be indicated by either different regression slopes or different intercepts between subgroups.

Study 2: Results

Descriptive Statistics

Descriptive statistics and correlations for all study variables are reported in Tables 7 and 8. I also examined the distribution of each of the variables of interest to assess normality of the data and identify any outliers. Overall, the distribution for age was slightly right-skewed such that a larger proportion of the sample was younger and the distributions of performance ratings were slightly left-skewed such that a larger proportion of the sample received higher ratings on both outcomes of performance. The test score distribution was highly left-skewed such that a majority of the sample responded to most of the items correctly. I conducted cube and square transformations of the test score data to attempt to mitigate the effects of the skewness; however, the transformed data did not produce different results for the statistical analyses described below. Thus, all results are based on the original data rather than the transformed data. Two outliers were identified in the data based on Grubbs' test for outliers (Grubbs, 1969)—one case in which a participant scored significantly lower on the test and one case in which a participant received a significantly lower performance rating for sales performance. These cases were removed from any analyses that included the respective variables.

Adverse Impact Analysis

I found small to moderate differences in test scores for the test of crystallized abilities between the subgroup below the age of 40 and those age 40 years and older, such that the older group attained lower scores on average (d = .36). Given that 47% of the total sample achieved the highest possible score on the test, adverse impact could not be evaluated in the extreme selection scenario. At a moderate overall selection ratio of 47%, adverse impact based on age was indicated with an adverse impact ratio of .62, indicating that those 40 years and older would have been selected at a rate that was 62% of the rate at which candidates under 40 years of age would have been selected. The differences in selection ratios were statistically significant based on Fisher's Exact test ($p < .01$) and the Z test of proportions ($Z = -3.45$). The results of a sensitivity analysis indicated that given the selection ratio of .37 for the older subgroup, the selection ratio of the younger subgroup (i.e., .60) was just below the minimum that would have allowed me to detect a statistically significant difference between proportions (i.e., .61). Additionally, tests of practical significance all met sufficient criteria to indicate adverse impact based on the phi value ($\varphi = .23$), the odds ratio ($OR = 2.52$; 95% CI = 1.48 – 4.29), and the h statistic ($h = .45$). See Table 9 for a comparison of adverse impact indicators for the test of crystallized abilities.

Linear Relations between Age and Cognitive Ability

I conducted a multiple linear regression analysis to examine linear relations between age and crystallized abilities, first regressing the test score on the control variables (i.e., race/ethnicity and gender) and then adding the predictor of interest (i.e., age at the time the test was taken). There were small, negative relations between age and test score ($\beta = -.26$, $p < .01$) when controlling for race/ethnicity and gender. Additionally, adding age to the regression model significantly increased the R^2 value compared to the original model with just the control variables. See Table 10 for a summary of the regression results. The results of a sensitivity analysis showed that the observed effect size ($f^2 = .80$) far exceeded the smallest effect that could be detected given the sample size, alpha level, and statistical power ($f^2 = .06$).

Differential Validity

The correlation between the test score and overall performance was $r = .28$ for the younger subgroup and $r = .08$ for the older subgroup. Using the Fisher r-to-Z transformation, I calculated the statistical significance of the difference in correlations and found that the difference was nonsignificant ($p = .12$). I also calculated Cohen's q to interpret the effect size of the difference and found a small effect ($q = .22$). However the results of a sensitivity analysis indicated that the smallest effect that could be detected given the sample sizes, alpha level, and statistical power was .48. I found similar results when using the sales-specific performance measure, as the test-criterion correlation was $r = .23$ for the younger subgroup and $r = .03$ for the older subgroup. The difference was not statistically significant ($p = .06$), but the effect size showed that there was a small effect ($q = .20$).

I also conducted the corrections for first range restriction and then unreliability. The corrected coefficients for range restriction were the same as the uncorrected coefficients; however, the corrected coefficients for unreliability were stronger as expected for the measure of overall job performance ($r = .37$ for younger; $r = .10$ for older) as well as for the measure of sales performance ($r = .30$ for younger; $r = .04$ for older).

Differential Prediction

Regarding differential prediction, comparisons of the regression equations for each subgroup are presented in Figures 4 and 5. There were small differences in both the slope ($\Delta = .14$) and intercept ($\Delta = 1.76$) for the overall performance outcome such that the slope was larger for the younger subgroup and the intercept was larger for the older subgroup. Similarly, there were differences in the slope ($\Delta = .23$) and intercept ($\Delta =.22$) for the sales performance outcome such that the slope was larger for the younger subgroup and the intercept was larger for the older subgroup. Results of the multiple linear regression analysis in which each performance outcome was regressed first on the test score, then on the age subgroup variable, and finally on the interaction between test score and subgroup indicated that the interaction term was not statistically significant for overall performance ($p = .07$) nor for sales performance ($p = .10$; Table 11).

Study 2: Discussion

The results of Study 2 indicated that there were small differences in crystallized cognitive abilities such that participants under the age of 40 scored .36 standard deviation units higher on the test than participants who were age 40 years or older. Hypothesis 1 regarding age-based adverse impact was not supported, as adverse impact was present at the 47% overall selection ratio according to the adverse impact ratio, both tests of statistical significance, and all three indicators of practical significance. Whereas 60% of those under the age of 40 responded to all 13 items correctly, only 37% of those age 40 years and older achieved the same maximum score. Additionally, I found a small, negative relation between age and test performance for the crystallized ability measure. These results contradict previous literature and theory which has suggested that *Gc* continues to increase as individuals age (Kanfer & Ackerman, 2004; Salthouse, 2009). These results as well as their implications are discussed further below.

Regarding differential validity, the results provided mixed support for Hypothesis 3, as differences in test-criterion correlation coefficients between the two age subgroups were not significant for the overall performance criterion nor for the sales-specific performance criterion, but the effect sizes showed small effects for the differences between subgroups in the opposite direction as hypothesized. Thus, the test of crystallized abilities was more strongly related to both overall performance and sales-specific performance for the younger subgroup compared to the older subgroup. However, Hypothesis 4 was supported, as the nonsignificant test-subgroup interaction terms indicated that differential prediction was not present.

There are several potential explanations for the findings suggesting that meaningful age-related differences existed in terms of both mean test score and consequent adverse impact as well as validity coefficients. First, although the test was not timed, it is possible that participants felt self-imposed time pressure given that they completed the test during work hours. If this was the case, processing speed theory (Salthouse, 1996) would suggest that younger participants would be more likely than older participants to answer the items correctly in a short amount of time, given the negative relation between age and processing speed. Additionally, given the design of the task intended to measure crystallized abilities, the measure itself may have been confounded with indicators of fluid ability including working memory and numerical reasoning. For example, the test required participants to integrate information from several sales reports which could only be accessed one at a time. Thus, in order to correctly answer some of the items, participants needed to retain the information from each relevant report for long enough to synthesize the information and arrive at the answer. Further, although the test items were much more domain-specific compared to general tests of numerical reasoning, many items required some level of numerical or mathematical ability which could be considered indicative of fluid cognitive abilities. Finally, although these findings were contrary to previous empirical studies regarding age differences in *Gc*, the measure of *Gc* used in this study was markedly different from those used previously, as many previous studies have used measures of verbal ability or specific knowledge to assess *Gc* (e.g., Klein et al., 2015; Ones et al., 2017).

Regarding the differences in the validity coefficients such that the test was more strongly related to overall performance and sales-specific performance for younger employees, it is important to note that the test reflected only one specific task that is involved in participants' particular roles. That is, although it is necessary for employees to be able to read and interpret the types of sales reports included in the test, it is possible that as one gains more experience in the role other aspects of the job (e.g., maintaining client relationships) become more important to success. Given that older employees are more likely to have more experience in their particular roles, the results suggest that other job-related tasks contribute to high performance ratings despite lower performance on this particular task. Additionally, following SOC theory (Baltes & Baltes, 1990), it is possible that older employees who have recognized declines in their ability to interpret sales reports such as those included in the test have found other ways to compensate for declines in their abilities—perhaps by delegating this task to subordinates or relying on other methods of digesting this information.

General Discussion

The purpose of the studies described above was to examine age-related differences in selection tests of both fluid and crystallized cognitive abilities in relation to test performance as well as criterion-related validity. Although the different samples used for each study prevents an exact comparison of results, the similarities between samples and procedures suggests that a consideration of the results in parallel is warranted. Both samples included similar numbers of incumbents working at the same organization and from the same range of management levels who were randomly selected to participate in concurrent validation studies. Additionally, performance data was collected in a similar manner, involving supervisory ratings.

The results of the current studies indicated that cognitive ability testing for selection purposes has the potential to lead to age-based adverse impact. Contrary to hypotheses, this was true for both tests of fluid cognitive abilities as well as a test of crystallized cognitive abilities. The results of Study 1 suggested that measures of deductive reasoning may be the most susceptible to age-based adverse impact based on adverse impact ratios and indicators of practical significance. The results of Study 2 also indicated that age-based adverse impact was likely to be present given a moderate overall selection ratio However, although age was negatively related to performance on all three tests of fluid abilities in Study 1 as well as the test of crystallized abilities in Study 2, I did not find evidence of differential validity or differential prediction for the numerical or deductive reasoning tests in Study 1. The figural reasoning test resulted in differential validity such that that there was a stronger test-criterion relation for the older subgroup compared to the younger subgroup. Conversely, in Study 2 the test of crystallized abilities was more strongly related to performance for the younger subgroup compared to the older subgroup, but there was no evidence of differential prediction. Thus, whereas the results of both studies indicated the potential for age-based adverse impact and showed negative linear relations between age and test performance across measures of both fluid and crystallized cognitive abilities, neither study indicated differential prediction between age subgroups.

Theoretical Implications

The findings of these studies have important implications for existing theory. First, although age-related declines in *Gf* and increases in *Gc* are well-documented in the cognitive literature, few studies (with the exception of Klein et al., 2015) have focused on how these age-related changes manifest in an organizational selection context. Although several of my hypotheses were not supported, the findings reported contribute to the general understanding of the relation between age and performance on cognitive ability tests for selection purposes. Additionally, a major strength of my studies was the inclusion of job performance data that allowed me to assess the extent to which measures of *Gf* and *Gc* differentially predict job performance for younger and older employees. Previous theoretical work related to aging (e.g., SOC theory, Baltes & Baltes, 1990) suggests that older employees may be able to use their previous experience and higher levels of *Gc* to compensate for age-related declines in *Gf*, ultimately leading to similar levels of job performance when compared to younger employees. Further, the theoretical link between cognitive ability and job performance is based on evidence that employees with high levels of *Gf* are able to acquire more knowledge at a faster rate once they are on the job (Borman et al., 1991; Hunter, 1986; McCloy et al., 1994; Schmidt et al., 1986). These theories point to the premise that tests of *Gf* may not predict job performance as well for experienced job candidates who already possess relevant job knowledge. However, no previous studies have empirically tested for differential validity and differential prediction of cognitive ability tests based on age (Fisher et al., 2017; Klein et al., 2015). Although the results of my studies provide mixed initial evidence regarding differential validity and prediction based on age, they ultimately point to the fact that questions related to these theoretical inconsistencies remain.

Additionally, the results highlight the importance of examining measures of narrow abilities (e.g., numerical reasoning, deductive reasoning, and figural reasoning) separately rather than generalizing at the broader level (i.e., *Gf* and *Gc*) when it comes to mean subgroup differences and differential validity and prediction. In the current studies, conclusions related to adverse impact and differential validity and prediction differed according to the particular measure of *Gf*. For example, indicators of practical significance pointed to meaningful subgroup differences in test performance for the deductive reasoning section but not for the numerical reasoning or the figural reasoning sections. Additionally, whereas previous studies have consistently shown that older individuals perform better on tests of verbal ability, the current studies showed that younger individuals performed better on tests of a domain-specific skill. Thus, these findings call for more focus on the differential implications of selection tests of narrow cognitive abilities.

Practical Implications

The findings also have several important implications for practice. First, the fact that adverse impact based on age was indicated at least to some extent for all four measures of cognitive ability included across the two studies provides further evidence of the heavily-supported notion that organizations should avoid using cognitive ability tests alone for employee selection (e.g., Outtz, 2011; Schmidt et al., 2016). Given the time and financial costs of discrimination suits, organizations are well-advised to avoid using tests that result in adverse impact for protected subgroups, even if they are able to provide evidence for the criterion-related validity of the test, thus making use of these tests legally defensible. Beyond the legal implications of discriminating against a protected class (i.e., those age 40 years and older), lower pass rates for older applicants could lead to a less diverse employee population, causing organizations to miss out on the benefits of age diversity on both the individual and firm levels (Konrad, 2003; Kunze et al, 2011; Rynes et al.,1997). To avoid this, selection systems should include additional assessments that are associated with fewer subgroup mean differences, such as personality tests (Outtz, 2011).

Additionally, the results of the exploratory analyses in Study 1, wherein the cutoff for determining the age subgroups was adjusted to 45, suggests that there are important practical implications for age diversity regarding tests of fluid abilities beyond any legal implications. Although the subgroup sample sizes were small, the effect sizes indicated that there were moderate differences in test performance between those under 45 and those 45 years and older, particularly for the numerical and figural reasoning sections of the test. Additionally, practical indicators suggested meaningful differences in selection ratios between the two subgroups under most overall selection scenarios. This may not have legal implications given that the cutoff for the legally protected subgroup is 40, however, it suggests that job candidates aged 45 years or older would be less likely to be selected into the organization on the basis of the cognitive ability test scores. Given the potential relevant job experience and skills that a 45 year-old candidate could possess, the test could be a major barrier for an organization strategically aiming to hire an experienced individual for a specific job. Further, eventually the use of the test could severely limit the age diversity of the organization with a preference for younger employees.

Finally, organizations should also consider age differences in applicant reactions to selection tests, including cognitive ability tests. Test anxiety, self-efficacy related to test-taking, and attitudes toward testing can each influence test performance through cognitive interference processes (Ryan & Chan, 1999). A limited amount of research has explored these reactions related to age. For example, in a study of job applicants, Lounsbury, Bobrow, and Jensen (1989) reported that older applicants had more negative attitudes toward selection tests. Additionally, Ackerman, Beier, and Bowen (2002) explored age differences in individuals' views of their own intellectual abilities and knowledge. They found that self-assessments of verbal ability, a measure of crystallized abilities, were positively related to age, whereas individuals' views of their own math and spatial abilities, aspects of fluid abilities, were negatively related to age. Thus, older job applicants may be more likely to experience negative reactions to cognitive ability tests, and especially to those that do not seem relevant for the job for which they are applying. Although I am not aware of any empirical studies that have investigated this, it is plausible that older applicants may avoid applying to jobs at organizations that are known to use cognitive ability tests or that they may choose to drop out of the application process when they are asked to take a cognitive ability test. In this case, the potential for age discrimination and limited age diversity would be introduced before the testing process even began.

Limitations

There are several limitations of the two studies I conducted that should be taken into consideration when interpreting the results. First, although the implications of the study are most applicable for job candidates taking tests for selection purposes, both studies used samples of incumbents who completed the tests as part of a concurrent validation study. Several prior studies have examined differences in test motivation that may lead to faking behaviors on tests such as personality assessments such that applicants tend to indicate more favorable responses than incumbents (e.g., Hough, 1998; Robie, Zickar, & Schmit, 2001). Although faking is not a concern for cognitive ability tests, it is likely that differences in test motivation would influence test performance such that incumbents for whom the test had no consequences may not have put as much time or thought into their responses as applicants who are motivated to receive a job offer would. Additionally, because all participants in these two studies had already been selected by the organization and therefore already met a threshold of cognitive ability assessed by the preexisting selection tests, there was likely range restriction on cognitive ability test scores. Although I corrected the validity coefficients for range restriction and found that they did not differ from the uncorrected coefficients, the corrections were conducted under the assumption of direct restriction based on the predictors. In reality, the restriction more likely occurred as a function of the previous selection system under which the participants were selected into the organization rather than on the predictors examined in the studies. In order to more accurately correct for indirect range restriction on a related variable, the unrestricted variance of the other variable (i.e., previous selection tests) must be known. Thus, although it was not possible to correct for this type of range restriction in the current studies, future studies should take this into account whenever possible.

Another limitation is the fact that my adverse impact analyses were based on the assumption that the tests were used in a single-hurdle selection context. This assumption does not reflect the reality of the selection system used by the organization, as several other assessments (e.g., cultural fit assessments, behavioral interviews) are used in conjunction with cognitive ability assessments. In fact, the literature strongly advises against using cognitive ability tests alone to select employees, given the well-documented evidence related to subgroup differences and consequential adverse impact (e.g., Outtz, 2011).

Finally, there were limitations associated with the use of supervisory ratings as a measure of job performance. Although any influences of subjective bias were removed from the selection system by focusing on an objective measure of cognitive ability, age-related biases can certainly influence the way supervisors rate their employees' performance. The results of several previous studies indicate that negative attitudes and cognitions about older workers lead to lower supervisory performance ratings (Cleveland & Landy, 1981; Gordon & Arvey, 2004; Bal et al., 2011). Further, the results of Waldman and Avolio's (1994) meta-analysis showed that whereas supervisory ratings tended to be lower for older employees, measures of individual productivity showed age-related increased in job performance. Thus, the use of supervisory performance ratings as the criterion in this study may introduce errors resulting from subjective biases.

Future Research

There are several directions for future research related to the topic of age discrimination associated with cognitive ability testing. First, future studies can address the limitations described above by using applicant samples instead of incumbents, evaluating age-based adverse impact when cognitive ability tests are used in multiple-hurdle or compensatory selection systems, and using other criterion measures such as productivity measures or peer ratings. Additionally, studies of within-person differences would allow for a more direct comparison of differential validity of tests of *Gf* and *Gc*. That is, studies in which participants take both types of tests within similar timeframes would allow researchers to assess the relation between performance on each test and a single criterion measure. Finally, longitudinal studies in which a sample of participants was given the same tests of *Gf* and *Gc* at different points in time across several years would provide additional insight regarding age-related increases or declines in cognitive ability test performance, although practice effects would need to be accounted for.

Conclusion

Given the aging workforce and increasing patterns of job mobility, age discrimination in the context of organizational selection continues to be an important topic. Previous literature has not addressed important research questions regarding the differential validity and prediction of cognitive ability tests based on age, and few studies have examined age-based adverse impact or continuous relations between cognitive ability selection test performance and age. These studies represent an important contribution to the literature regarding cognitive ability testing for selection and age discrimination. The results provide initial insight as to the nature of the relations between age and tests of both fluid and crystallized cognitive abilities for selection purposes. Finally, these studies emphasize the need for further research to continue to understand age-related issues in cognitive ability testing.

Table 1.

Study 1- Sample demographic characteristics.

Variable	N	%
Gender		
Female	91	42.5
Male	123	57.5
Race/Ethnicity		
Caucasian/White	120	56.1
Asian/Pacific-Islander	33	15.4
African-American/Black	25	11.7
Hispanic/Latino	30	14.0
Missing	6	2.8
Age		
Under 40 years	135	63.1
40 years and older	79	36.9

Table 2.
Descriptive statistics and correlations for Study 1.

	M	*SD*	1	2	3	4	5	6	7	8	9
1. Age (in years)	38.01	7.98	-								
2. Gender	0.43	.50	-.11	-							
3. White	.56	.50	.06	-.08	-						
4. Black	.12	.32	-.01	.04	-.41**	-					
5. Hispanic	.14	.35	-.02	-.08	-.46**	-.15*	-				
6. Asian	.15	.36	-.04	.08	-.48**	-.16*	-.17*	-			
7. Numerical Score	12.43	2.37	-.26**	-.18**	.02	-.19**	-.07	.20**	-		
8. Logic Score	7.36	1.89	-.21**	-.11	.25**	-.09	-.24**	.00	.36**	-	
9. Figural Score	9.70	3.09	-.25**	-.18**	.25**	-.19**	-.26**	.11	.45**	.40**	
10. Performance	5.05	.88	-.06	-.04	.04	-.08	.02	-.04	.14*	.14*	.06

Note. N = 214 for all variables. Demographic control variables were dummy-coded (i.e., male compared to female, White compared to non-White, etc.). *Correlation is significant at the $p < .05$ level. **Correlation is significant at the $p < .01$ level.

Table 3.
Indicators of age-based adverse impact for Study 1.

			Statistical Significance		Practical Significance Indicators			
	d	AI Ratio	Fisher's Exact (p value)	Z_{IR}	Phi	Odds Ratio	OR 95% CI	*h*
Numerical (21% SR)	.45	**.79**	.49	-.81	.05	1.33	.65 - 2.69	.11
Deductive (12% SR)	.45	**.51**	.13	-1.75	**.11**	2.12	.81 - 5.52	**.23**
Figural (9% SR)	.46	**.61**	.46	-1.09	.07	1.71	.59 - 4.95	.15

Note. Bolded numbers indicate empirical evidence of adverse impact.

Table 4.
Indicators of age-based adverse impact for Study 1 exploratory analyses.

			Statistical Significance		Practical Significance Indicators			
	d	AI Ratio	Fisher's Exact (p value)	Z_{IR}	Phi	Odds Ratio	OR 95% CI	*h*
Numerical (21% SR)	.66	**.42**	.08	**-2.53**	**.13**	2.78	.93 - 8.30	**.37**
Deductive (12% SR)	.35	1.00	1.00	-.01	.00	.99	.35 - 2.81	.00
Figural (44% SR)	.54	**.73**	.22	-1.55	**.10**	1.67	.81 - 3.38	**.25**

Note. Bolded numbers indicate empirical evidence of adverse impact.

Table 5.
Regression results for test section scores on demographic variables and age for Study 1.

	Model 1			Model 2		
	Numerical	*Deductive*	*Figural*	*Numerical*	*Deductive*	*Figural*
Female	-.19**	-.11	-.18**	-.22**	-.14*	-.21**
White	-.06	.29	.22	-.06	-.29	.22
Asian	.14	.14	.19	.13	.13	.18
Black	-.20	.05	-.09	-.21	.04	-.09
Hispanic	-.11	-.09	-.15	-.13	-.10	-.17
Age				-.28**	-.24**	-.29
R^2	.11	.10	.17	.18	.16	.25**
ΔR^2				.08	.06	.08

Note. Demographic control variables were dummy-coded (i.e., female compared to male, White compared to non-White, etc.). *Coefficient is significant at the p < .05 level. **Coefficient is significant at the p < .01 level.

Table 6.
Regression results for differential prediction analyses for Study 1.

	Model 1			Model 2			Model 3		
	Numerical	*Deductive*	*Figural*	*Numerical*	*Deductive*	*Figural*	*Numerical*	*Deductive*	*Figural*
Test Score	.14	.14	.06	.13	.14	.05	.09	.21	-.22
Age Subgroup				-.02	-.02	-.04	-.09	.07	-.32
Test Score*Age							.07	-.11	.35
R^2	.02	.02	.00	.02	.02	.01	.02	.02	.01
ΔR^2				.00	.00	.01	.00	.00	.00

Note. All coefficients are nonsignificant.

Table 7.

Study 2- Sample demographic characteristics.

Variable	N	%
Gender		
Female	115	49.6
Male	117	50.4
Race/Ethnicity		
Caucasian/White	161	69.4
Asian/Pacific-Islander	20	8.6
African-American/Black	21	9.1
Hispanic/Latino	24	10.3
Missing	6	2.8
Age		
Under 40 years	107	46.1
40 years and older	125	53.9

Table 8.

Descriptive statistics and correlations for Study 2.

	M	*SD*	1	2	3	4	5	6	7	8
1. Age (in years)	42.05	11.65	-							
2. Gender	.50	.50	-	-						
3. White	.71	.46	-	-.07	-					
4. Black	.09	.29	-	.11	-.49**	-				
5. Hispanic	.11	.31	-	-.06	-.53**	-.11	-			
6. Asian	.09	.28	-	.09	-.45**	-.10	-.11	-		
7. *Gc* Score	12.03	1.29	-.26**	.08	.01	-.05	-.08	.11	-	
8. Overall Performance	5.12	.75	.04	.07	.08	-.06	-.06	.03	.15*	-
9. Sales Performance	5.03	.95	.07	.09	.15*	-.09	-.12	.04	.09	.84**

Note. $N = 232$ for all variables. Demographic control variables were dummy-coded (i.e., male compared to female, White compared to non-White, etc.). *Correlation is significant at the $p < .05$ level. **Correlation is significant at the $p < .01$.

Table 9.
Indicators of age-based adverse impact for Study 2.

			Statistical Significance		Practical Significance Indicators			
	d	AI Ratio	Fisher's Exact (p value)	Z_{IR}	Phi	Odds Ratio	OR 95% CI	*h*
Gc (47% SR)	.39	**.62**	**< .001**	**-3.45**	**.23**	**2.52**	**1.48 - 4.29**	**.45**

Note. Bolded numbers indicate empirical evidence of adverse impact.

Table 10.
Regression results for test score on demographic variables and age for Study 2

	Model 1	Model 2
Male	-.09	-.04
White	.02	.09
Age		$-.27^{**}$
R^2	.01	.07
ΔR^2		.06

Note. **Coefficient is significant at the $p < .01$.

Table 11.

Regression results for differential prediction analyses for Study 2.

	Model 1		Model 2		Model 3	
	Overall Performance	*Sales Performance*	*Overall Performance*	*Sales Performance*	*Overall Performance*	*Sales Performance*
Test Score	.15*	.09	.16*	.11	.32**	.26*
Age Subgroup			.07	.11	1.24	1.21
Test Score*Age					-1.12	-1.09
R^2	.02	.01	.03	.02	.04	.03
ΔR^2			.01	.01	.01	.01

Note. *Coefficient is significant at the $p < .05$ level. **Coefficient is significant at the $p < .01$ level.

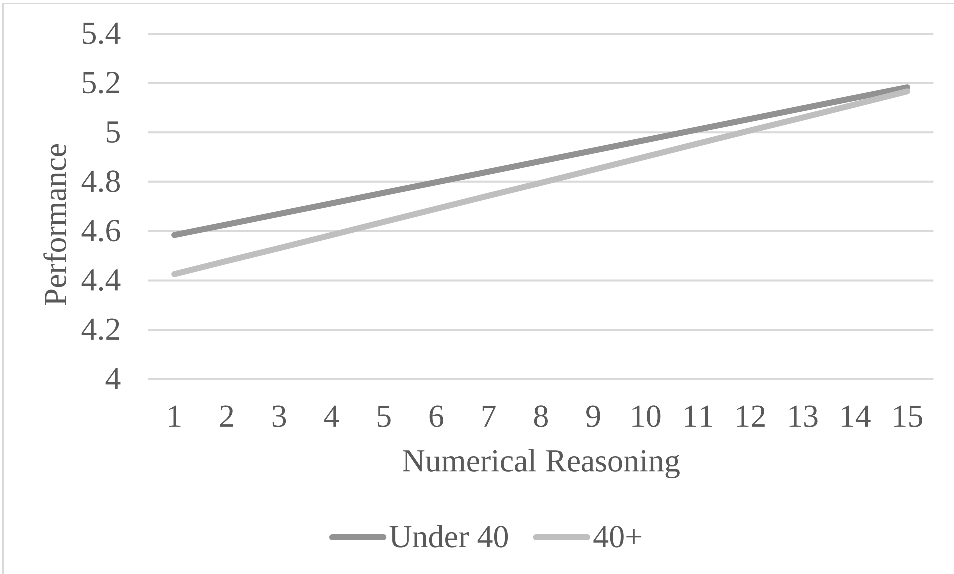

Figure 1. Performance regressed on numerical reasoning test scores for both age subgroups

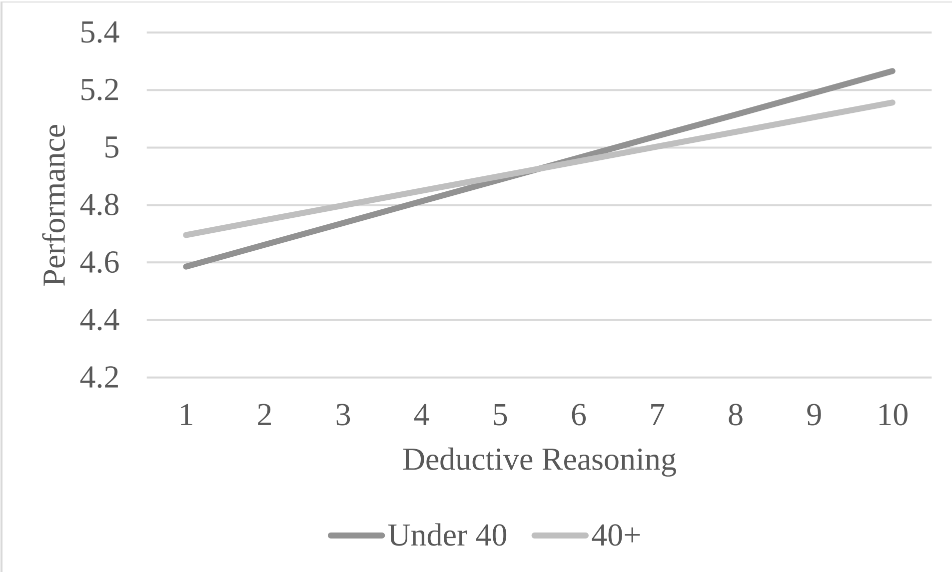

Figure 2. Performance regressed on deductive reasoning test scores for both age subgroups

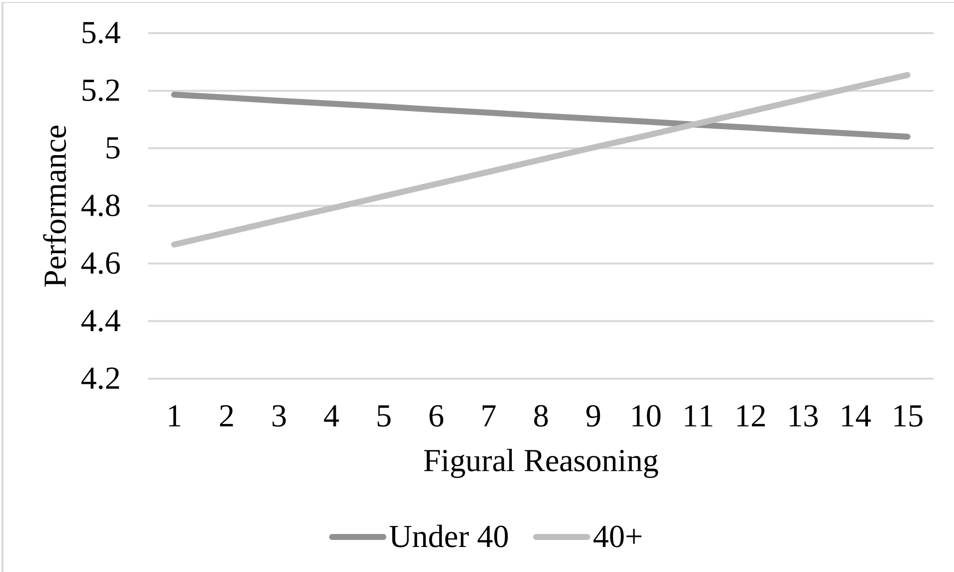

Figure 3. Performance regressed on figural reasoning test scores for both age subgroups.

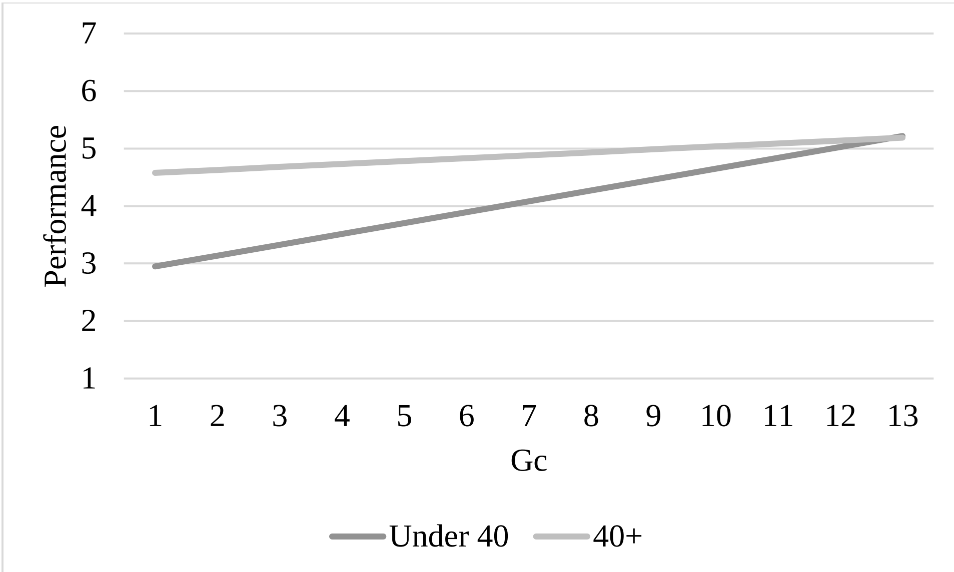

Figure 4. Overall performance regressed on *Gc* test scores for both age subgroups.

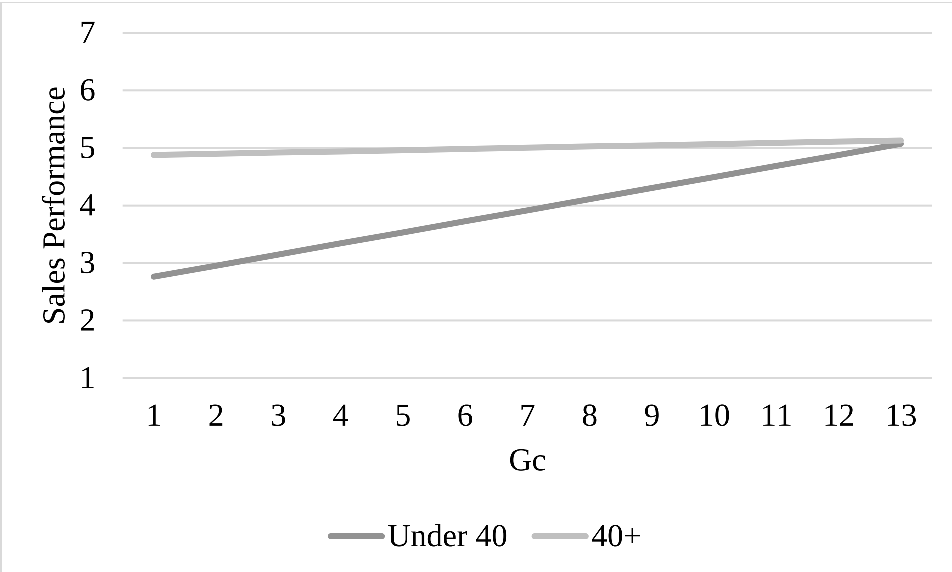

Figure 5. Sales performance regressed on *Gc* test scores for both age subgroups.

References

Albemarle Paper Co. v. Moody (1975). 422 U.S. 405.

Alexander, R. A., Alliger, G. M., & Hanges, P. J. (1984). Correcting for range restriction when the population variance is unknown. *Applied Psychological Measurement, 8*(4), 431-437.

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 185-202). New York: Guilford Press.

Almeida, P., Dokko, G., & Rosenkopf, L. (2003). Startup size and the mechanisms of external learning: increasing opportunity and decreasing ability? *Research Policy*, *32*(2), 301-315.

Avolio, B. J., & Waldman, D. A. (1994). Variations in cognitive, perceptual, and psychomotor abilities across the working life span: Examining the effects of race, sex, experience, education, and occupational type. *Psychology and Aging*, *9*(3), 430-442.

Bal, A. C., Reiss, A. E. B., Rudolph, C. W., & Baltes, B. B. (2011). Examining positive and negative perceptions of older workers: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *66B*(6), 687-698.

Baltes, P. B., & Baltes, M. M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. In P. B. Baltes, & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 1-34). New York: Cambridge University Press.

Beier, M. E., Young, C. K., & Villado, A. J. (2017). Job knowledge: Its definition, development, and measurement. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work & organizational psychology* (pp. 279-298). Los Angeles: SAGE.

Berry, C. M. (2015). Differential validity and differential prediction of cognitive ability tests: Understanding test bias in the employment context. *Annual Review of Organizational Psychology and Organizational Behavior, 2*(1), 435-463.

Beus, J. M., Jarrett, S. M., Taylor, A. B., & Wiese, C. W. (2014). Adjusting to new work teams: Testing work experience as a multidimensional resource for newcomers. *Journal of Organizational Behavior, 35*(4), 489-506.

Beyer, J. M., & Hannah, D. R. (2002). Building on the past: Enacting established personal identities in a new work setting. *Organization Science, 13*(6), 636-652.

Biddle, D. A., & Morris, S. B. (2011). Using Lancaster's mid-P correction to the Fisher's exact test for adverse impact analyses. *Journal of Applied Psychology, 96*(5), 956-965.

Birren, J. A. (1969). Age and decision strategies. In A. T. Welford & J. A. Birren (Eds.), *Interdisciplinary topics in gerontology (Vol. 4)* (pp. 23-36). New York: Karger.

Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on Black–White mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology, 66*(1), 91-126.

Boehm, S. A., Kunze, F., & Bruch, H. (2014). Spotlight on age-diversity climate: The impact of age-inclusive HR practices on firm-level outcomes. *Personnel Psychology, 67*(3), 667-704.

Bonnet, D. G. & Price, R. M. (2007). Statistical inference for generalized Yule coefficients in 2x2 contingency tables. *Sociological Methods & Research, 35*, 429-446.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76*(6), 863-872.

Bravery, K., Silva, J., Jones, K., Shellenback, K., Polovina, S., Mishra, P., Hudson, R., Taylor, M., Harley, G., de Maio, A. (2017). Talent trends: 2017 global study. New York: Mercer.

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: a developmental cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience, 7*, 1-13.

Bryan, J. & Luszcz, M. A. (1996). Speed of information processing as a mediator between age and free-recall performance. *Psychology and Aging, 11*(1), 3-9.

Byron, K. (2005). A meta-analytic review of work–family conflict and its antecedents. *Journal of Vocational Behavior*, *67*(2), 169-198.

Carless, S. A. (2009). Psychological testing for selection purposes: A guide to evidence-based practice for human resource professionals. *International Journal of Human Resource Management*, *20*(12), 2517-2532.

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge, UK: Cambridge University Press.

Cascio, W. F., & Aguinis, H. (2010). *Applied psychology in human resource management (7 edition)*. Upper Saddle River, NJ: Pearson.

Cascio, W. F., & Boudreau, J. W. (2011). Investing in people: Financial impact of human resource initiatives. Upper Saddle River, NJ: Pearson Education, Inc.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153-193.

Chamorro-Premuzic, T., & Furnham, A. (2010). *The psychology of personnel selection*. Cambridge, UK: Cambridge University Press.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance, 7*(2), 467-476.

Choudhry, N. K., Fletcher, R. H., Soumerai, S. B. (2005). Systematic review: The relationship between clinical experience and quality of health care. *Annals of Internal Medicine*, *142*, 260-273.

Civil Rights Act, 42 USCS § 2000e (1964).

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, *5*, 115-124.

Cleveland, J. N. & Landy, F. J. (1981). The influence of rater and ratee age on performance judgments. *Personnel Psychology, 34*(1), 19-29.

Cobb, B. B., Lay, C. D., & Bourdet, N. M. (1971). The relationship between chronological age and aptitude test measures of advanced-level air traffic control trainees. Oklahoma City, OK: Civil Aeromedical Institute.

Cohen, A. C. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics,* 7(3), 217-237.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* San Diego, CA: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Collins, M. W., & Morris, S. B. (2008). Testing for adverse impact when sample size is small. *Journal of Applied Psychology, 93*(2), 463-471.

Cornelius, S. W., & Caspi, A. (1987). Everyday problem solving in adulthood and old age. *Psychology and Aging*, *2*(2), 144-153.

Cuddy, A. J., Norton, M. I., & Fiske, S. T. (2005). This old stereotype: The pervasiveness and persistence of the elderly stereotype. *Journal of Social Issues*, *61*(2), 267-285.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168.

Dilchert, S. (2017). Cognitive ability. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work & organizational psychology* (pp. 248–276). Los Angeles, CA: SAGE.

Dokko, G., Wilk, S. L., & Rothbard, N. P. (2009). Unpacking prior experience: How career history affects job performance. *Organization Science*, *20*(1), 51-68.

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish mental survey. *Intelligence*, *28*, 49-55.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130-147.

Drasgow, F. & Kang, T. (1984). Statistical power of differential validity and differential prediction analyses for detecting measurement nonequivalence. *Journal of Applied Psychology*, *69*(3), 498-508.

Equal Employment Opportunity Commission (2017). Age Discrimination in Employment Act (ADEA) Charges. Retrieved from https://www.eeoc.gov/eeoc/statistics/enforcement/adea.cfm

Finkel, D., Reynolds, C. A., McArdle, J. J., & Pedersen, N. L. (2007). Age changes in processing speed as a leading indicator of cognitive aging. *Psychology and Aging*, *22*(3), 558-568.

Finkelstein, L. M., & Farrell, S. K. (2012). An expanded view of age bias in the workplace. In K. S. Shultz & G. A. Adams (Eds.), *Aging and work in the 21st century* (pp. 73–108). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Fisher, G. G., Truxillo, D. M., Finkelstein, L. M., & Wallace, L. E. (2017). Age discrimination: Potential for adverse impact and differential prediction related to age. *Human Resource Management Review*, *27*(2), 316–327.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation. Needham Heights, MA: Allyn & Bacon

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *95*, 29-51.

Flynn, J. R. (1998). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 25-66). Washington, DC: American Psychological Association.

Gagne, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist*, *39*(4), 377-385.

Gastwirth, J. L. (1988). *Statistical reasoning in law and public policy* (Vol. 1). San Diego, CA: Academic Press.

Ghiseli, E. E. (1964). *Theory of psychological measurement*. New York, NY: McGraw-Hill.

Gonzalez-Mulé, E., Mount, M. K., & Oh, I.-S. (2014). A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology*, *99*(6), 1222-1243.

Gordon, R. A., & Arvey, R. D. (2004). Age bias in laboratory and field settings: A meta-analytic investigation. *Journal of Applied Social Psychology*, *34*(3), 468-492.

Griggs v. Duke Power Co. (1971). 401 U.S. 424.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics, 11*, 1-21.

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed). New York, NY: Routledge.

Gutman, A., Koppes, L. L., & Vodanovich, S. J. (2010). *EEO law and personnel practices*. New York: Taylor and Francis Group, LLC.

Heil, M. C. (1999). An investigation of the relationship between chronological age and job performance for incumbent air traffic control specialists (DOT/FAA/AM-99/18). Oklahoma City, OK: FAA Civil Aeromedical Research Institute.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75*, 242-259.

Horn, J. L. (1975). Psychometric studies of aging and intelligence. *Psychopharmacology Bulletin*, *11*(4), 44-45.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 267-300). New York, NY: Wiley.

Horn, J. L. (1998). A basis for research on age differences in cognitive capabilities. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57-92). New York, NY: Routledge.

Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253-270.

Horn, J. L. & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 107-129.

Horn, J. L. & Donaldson, G. (1980). Cognitive development in adulthood. In O. G. Brim & J. Kagan (Eds.), *Constancy and change in human development* (pp. 445-529). Cambridge, MA: Harvard University Press.

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

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*(1-2), 152-194.

Hülsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, *15*(1), 3-18.

Hunter, J. E. (1983). *The dimensionality of the General Aptitude Test Battery (GATB) and the dominance of general factors over specific factors in the prediction of job performance for the U.S. Employment Service*. Washington, DC: U.S. Employment Service, U.S. Department of Labor.

Hunter, J. E. (1985). *Differential validity across jobs in the military*. Rockville, MD: Research Applications.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*(3), 340–362.

Hunter, J. E. (1989). *The Wonderlic Personnel Test as a predictor of training success and job performance*. Libertyville, IL: Wonderlic.

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

Hunter, J. E. & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2*, 447-472.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. Journal of Applied Psychology, 91, 594-612.

Jacobs, R., Murphy, K., & Silva, J. (2013). Unintended consequences of EEO enforcement policies: Being big is worse than being bad. *Journal of Business and Psychology*, *28*(4), 467-471.

Kanfer, R. & Ackerman, P.L. (2004). Adult development and work motivation. *The Academy of Management Review*, *29*(3), 440-458.

Klein, R. M., Dilchert, S., Ones, D. S., & Dages, K. D. (2015). Cognitive predictors and age-based adverse impact among business executives. *Journal of Applied Psychology, 100*(5), 1497-1510.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Konrad, A. M. (2003). Defining the domain of workplace diversity scholarship. *Group & Organization Management*, *28*(1), 4-17.

Krantowitz, T. M. (2014). *2014 Global Assessment Trends Report*. Surrey: CEB-SHL.

Kunze, F., Boehm, S. A., & Bruch, H. (2011). Age diversity, age discrimination climate and performance consequences—a cross organizational study. *Journal of Organizational Behavior, 32*(2), 264-290.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*(2), 460-468.

Lord, F, & Novick, M. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Lyons, S. T., Schweitzer, L., & Ng, E. S. W. (2015). How have careers changed? An investigation of changing career patterns across four generations. *Journal of Managerial Psychology*, *30*(1), 8-21.

Masunaga, H., & Horn, J. (2001). Expertise and age-related changes in components of intelligence. *Psychology and Aging, 16*(2), 293-311.

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology, 24*(2), 222-243.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*(4), 493-505.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, *60*(1), 63-91.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730-740.

McDaniel, M. A. & Nguyen, N. T. (2001). Situational judgement tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*(1/2), 103-113.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology*, *73*(2), 327-330.

McEvoy, G. M., & Cascio, W. F. (1989). Cumulative evidence of the relationship between employee age and job performance. *Journal of Applied Psychology*, *74*(1), 11-17.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1-10.

Morgeson, F. P., Reider, M. H., Campion, M. A., & Bull, R. A. (2008). Review of research on age discrimination in the employment interview. *Journal of Business and Psychology*, *22*(3), 223-232.

Morrison, R. F., & Brantner, T. M. (1992). What enhances or inhibits learning a new job? A basic career issue. *Journal of Applied Psychology, 77*(6), 926-940.

Motowildo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, *10*(2), 71-83.

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, *91*(4), 749-761.

Motowidlo, S. J., Packard, J. S., & Manning, M. R. (1986). Occupational stress: Its causes and consequences for job performance. *Journal of Applied Psychology, 71*(4), 618-629.

Murphy, K. R., Cronin, B. E., & Tam, A. P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organizations. *Journal of Applied Psychology*, *88*(4), 660-671.

Murphy, K. R. & Davidshofer, C. O. (1988). *Psychological testing: Principles and applications*. Upper Saddle River, NJ: Prentice Hall.

Murphy, K. R., & Jacobs, R. R. (2012). Using effect size measures to reform the determination of adverse impact in equal employment litigation. *Psychology, Public Policy, and Law, 18*(3), 477-499.

Naude, M. N. (2018). Effects of time since graduation on cognitive ability selection test performance. Unpublished manuscript.

Neumark, D. (2009). The Age Discrimination in Employment Act and the challenge of population aging. *Research on Aging, 31*(1), 41-68.

Ng, T. W., & Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology*, *93*(2), 392-423.

Ng, T. W. H., & Feldman, D. C. (2012). Evaluating six common stereotypes about older workers with meta-analytical data. *Personnel Psychology, 65*(4), 821–858.

North, M. S., & Fiske, S. T. (2012). An inconvenienced youth? Ageism and its potential intergenerational roots. *Psychological Bulletin*, *138*(5), 982-997.

O'Connor v. Consolidated Coin Caterers Corp. (1996). 517 U.S. 308.

Ones, D. S., Dilchert, S., Chockalingam, V., & Salgado, J. (2017). Cognitive ability: Measurement and validity for employee selection. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp. 251–276). New York, NY: Routledge.

Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). Cognitive abilities. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 179–224). New York, NY: Oxford University Press.

Oswald, F. L., Dunleavy, E. M., & Shaw, A. (2017). Measuring practical significance in adverse impact analysis. In S. B. Morris & E. M. Dunleavy (Eds.), *Adverse impact analysis: Understanding data, statistics, and risk* (pp. 92-112). New York, NY: Routledge.

Outtz, J. (2011). The unique origins of advancements in selection and personnel psychology. In S. Zedeck (Ed.), *APA handbook of industrial organizational psychology, Volume 2. Selecting and developing members*. Washington, DC: American Psychological Association.

Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. In D. C. Park & N. Schwarz (Eds.), *Cognitive aging: A primer*. New York, NY: Routledge.

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging*, *11*(4), 621-637.

Peeters, M. C., & van Emmerik, H. (2008). An introduction to the work and well-being of older workers: From managing threats to creating opportunities. *Journal of Managerial Psychology*, *23*(4), 353-363.

Perlmutter, M. (1988). Cognitive potential throughout life. In J. E. Birren & V. Bengtson (Eds.), *Emergent theories of aging* (pp. 247-268). New York, NY: Springer.

Ployhart, R. E. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management, 32*(6), 868-897.

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*(1), 153-172.

Posthuma, R. A., & Campion, M. A. (2009). Age stereotypes in the workplace: Common stereotypes, moderators, and future research directions. *Journal of Management, 35*(1), 158-188.

Posthlethwaite, B. E. (2011). Fluid ability, crystallized ability, and performance across multiple domains: A meta-analysis. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Rao, H., & Drazin, R. (2002). Overcoming resource constraints on product innovation by recruiting talent from rivals: A study of the mutual fund industry, 1986-94. *Academy of Management Journal*, *45*(3), 491-507.

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1-48.

Raz, N., Lindenberger, U., Ghisletta, P., Rodrigue, K. M., Kennedy, K. M., & Acker, J. D. (2008). Neuroanatomical correlates of fluid intelligence in healthy adults and persons with vascular risk factors. *Cerebral Cortex*, *18,* 718-726.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, *1*(3), 86-89.

Robie, C., Zickar, M. J., & Schmit, M. J. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. *Human Performance*, *14*(2), 187-207.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, *54*(2), 297-330.

Roth, P. L., Bobko, P., & Switzer III, F. S. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal of Applied Psychology*, *91*(3), 507-522.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*(1), 66-80.

Rupp, D. E., Vodanovich, S. J., & Credé, M. (2006). Age bias in the workplace: The impact of ageism and causal attributions. *Journal of Applied Social Psychology*, *36*(6), 1337-1364.

Rynes, S. L., Orlitzky, M. O., & Bretz, R. D. (1997). Experienced hiring versus college recruiting: Practices and emerging trends. *Personnel Psychology*, *50*(2), 309-339.

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology*, *49*(3), 549-572.

Sackett, P. R. & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, *85*(1), 112-118.

Salgado, J. F. (2017). Using ability tests in selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention* (pp. 113–150). Chichester, UK: John Wiley & Sons, Ltd.

Salgado, J. F., & Anderson, N. (2002). Cognitive and GMA testing in the European community: Issues and evidence. *Human Performance, 15*(1/2), 75-96.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*(3), 573-605.

Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology*, *113*, 345-371.

Salthouse, T. A. (1994). Age-related differences in basic cognitive processes: Implications for work. *Experimental Aging Research, 20,* 249-255.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403-428.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiological Aging, 30,* 507-514.

Salthouse, T. A. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology*, *63*, 201-226.

Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology and Aging*, *13*(3), 445-461.

Salthouse, T. A. & Madden, D. J. (2013). Information processing speed and aging. In J. DeLuca & J. H. Kalmar (Eds.), *Information processing speed in clinical populations* (pp. 221-242). New York, NY: Taylor & Francis.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49*(4), 304-313.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*(1-2), 187-210.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, *36*(10), 1128-1137.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*(3), 432-439.

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*(2), 262-274.

Schmidt, F. L., Oh, I. S. & Shaffer, J. A. (2016). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 100 years of research findings. Working paper. University of Iowa, Iowa City, IA, available at: www.researchgate.net/profile/Frank_Schmidt10/publication/309203898_The_Validity_and_Utility_of_Selection_Methods_in_Personnel_Psychology_Practical_and_Theoretical_Implications_of_100_Years_of_Research_Findings/links/58051e2608aef87fbf3bbb52.pdf

Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy of range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61*, 827-868.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carrol model of intelligence. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). New York, NY: Guilford.

Smith, K. C. & McDaniel, M. A. (1998, April). *Criterion and construct validity evidence for a situational judgment test.* Paper presented at the 13th Annual Convention of the Society for Industrial and Organizational Psychology, Dallas, TX.

Smith v. City of Jackson (2005). 544 U.S. 228.

Society for Industrial and Organizational Psychology (2003). *Principles for the Validation and Use of Personnel Selection Procedures*. Bowling Green, OH: SIOP.

Stanek, K. C. & Ones, D. S. (2017). Taxonomies and compendia of cognitive ability and personality constructs and measures relevant to industrial, work, and organizational psychology. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work & organizational psychology* (pp. 320–365). Los Angeles, CA: SAGE.

Staudinger, U. M., Cornelius, S. W., & Baltes, P. B. (1989). The aging of intelligence: Potential and limits. *The Annals of the American Academy of Political and Social Science, 503*(1), 43-59.

Sternberg, R. J., & Detterman, D. K. (Eds.). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). Practical intelligence: The nature and role of tacit knowledge in work and at school. *Mechanisms of Everyday Cognition*, 205-227.

Sturman, M. C. (2003). Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management, 29*(5), 609-640.

Switzer, F. S., Roth, P. L., & Rosopa, P. J. (2016). Power and errors in adverse impact indices. *Human Performance, 29*(3), 159–171.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York: Wiley.

Tillsely, C., & Taylor, P. (2001). Managing the third age workforce: A review and agenda for research. In I. Golver & M. Branine (Eds.), *Ageism in work and employment* (pp. 311-326). Burlington, VT: Ashgate Publishing.

Tracey, T. J., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal?. *American Psychologist*, *69*(3), 218-229.

Trites, D. K. & Cobb, B. B. (1963). *Problems in air traffic management IV: Comparison of pre-employment, job-related experience, with aptitude tests as predictors of training and job performance as air traffic control specialists*. Oklahoma City, OK: FAA Civil Aeromedical Institute.

Trites, D. K. & Cobb, B. B. (1964). *CARI research on air traffic control specialists: Age, aptitude, and experience as predictors of performance*. Oklahoma City, OK: FAA Civil Aeromedical Institute.

Truxillo, D. M., Cadiz, D. M., & Rineer, J. R. (2014). The aging workforce: Implications for human resource management research and practice. In S. Jackson (Ed.), *The aging workforce: Implications for human resource management research and practice*. Oxford Handbooks Online: Business & Management.

Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*, *9*(4), 408-426.

Uniform Guidelines on Employee Selection Procedures (1978). 29 C.F.R. 1607.

Uppal, N., Mishra, S. K., & Vohra, N. (2014). Prior related work experience and job performance: Role of personality. *International Journal of Selection and Assessment, 22*(1), 39-51.

U.S. Bureau of Labor Statistics (2015). Civilian labor force by age, sex, race, and ethnicity, 1994, 2004, 2014, and projected 2024. Retrieved from https://www.bls.gov/news.release/ecopro.t01.htm.

U.S. Bureau of Labor Statistics (2017). Number of jobs, labor market experience, and earnings growth among Americans at 50: Results from a longitudinal survey. Retrieved from https://www.bls.gov/news.release/pdf/nlsoy.pdf.

Verhaeghen, P. & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122*(2), 231-249.

Vinchur, A. J., & Koppes, L. L. (2012). A history of personnel selection and assessment. In N. Schmitt (Ed.), *The Oxford Handbook of personnel assessment and selection* (pp. 9–30). New York: Oxford University Press.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, *8*(4), 216-226.

Waldman, D. A. & Avolio, B. J. (1986). A meta-analysis of age differences in job performance. *Journal of Applied Psychology, 71*, 33-38.

Waldman, D. A., & Avolio, B. J. (1993). Aging and work performance in perspective: Contextual and developmental considerations. *Research in Personnel and Human Resources Management*, *11*, 133-162.

Wanberg, C. R., Kanfer, R., Hamann, D. J., & Zhang, Z. (2016). Age and reemployment success after job loss: An integrative model and meta-analysis. *Psychological Bulletin*, *142*(4), 400-426.

Wang, M., Olson, D. A., & Shultz, K. S. (2012). *Mid and late career issues: An integrative perspective*. New York, NY: Routledge.

Warr, P., & Pennington, J. (1994). Occupational age-grading: Jobs for older and younger nonmanagerial employees. *Journal of Vocational Behavior, 45*(3), 328-346.

Wee, S., Newman, D. A., & Joseph, D. L. (2014). More than g: Selection quality and adverse impact implications of considering second-stratum cognitive abilities. *Journal of Applied Psychology*, *99*(4), 547-563.

Zedeck, S. (2010). Adverse impact: History and evolution. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 3-27). New York, NY: Routledge.

Zimprich, D., & Martin, M. (2002). Can longitudinal changes in processing speed explain longitudinal age changes in fluid intelligence? *Psychology and Aging*, *17*(4), 690-695.