

Spring 2016

# A Review of Court Cases Involving Cognitive Ability Testing and Employment Practices: 1992-2015

Nicholas H. Morris

Follow this and additional works at: <http://digitalcommons.wku.edu/theses>

 Part of the [Applied Behavior Analysis Commons](#), [Community Psychology Commons](#), and the [Industrial and Organizational Psychology Commons](#)

---

## Recommended Citation

Morris, Nicholas H., "A Review of Court Cases Involving Cognitive Ability Testing and Employment Practices: 1992-2015" (2016). *Masters Theses & Specialist Projects*. Paper 1575.  
<http://digitalcommons.wku.edu/theses/1575>

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact [topscholar@wku.edu](mailto:topscholar@wku.edu).

A REVIEW OF COURT CASES INVOLVING COGNITIVE ABILITY TESTING  
AND EMPLOYMENT PRACTICES: 1992-2015

A Thesis  
Presented to  
The Faculty of the Department of Psychological Sciences  
Bowling Green, Kentucky

In Partial Fulfillment  
Of the Requirements of the Degree  
Master of Sciences

By  
Nicholas Morris

May 2016

A REVIEW OF COURT CASES INVOLVING COGNITIVE ABILITY TESTING  
AND EMPLOYMENT PRACTICES: 1992-2015

Date Recommended April 21, 2016

Elizabeth L. Shoenfelt  
Elizabeth L. Shoenfelt, Director of Thesis

R. D. Brown  
Reagan D. Brown

Amber Schroeder  
Amber N. Schroeder

Liya 4/25/16  
Dean, Graduate School      Date

I dedicate this thesis to my parents, Richard and Donna Morris. Without them, I would have undoubtedly fallen short of my goals. Their love and encouragement is what sustains me and lifts me higher than I ever would have gone alone.

## ACKNOWLEDGEMENTS

I acknowledge my professors, whom I feel very lucky to have encountered in my graduate school career. Not only did they see this project to its completion, they have all helped me grow tremendously in areas where I was weak professionally and they could not have done so without having a genuine interest in my success as well as the success of my fellow students. I cannot speak highly enough of that. In particular, Betsy Shoenfelt has been a huge agent of change for me through her belief in me and providing me with the tools and experience I needed and will need to continually improve.

## CONTENTS

Introduction.....	1
General Mental Ability/Cognitive Ability .....	2
Cognitive Ability as a Predictor of Job Performance.....	2
Disagreement of General Cognitive Ability as a Predictor of Performance .....	4
The Risk of Using Cognitive Ability Tests .....	5
Should an Organization Use Cognitive Ability Testing?.....	6
Important Court Cases and Legislation .....	8
Reducing Risk of Adverse Impact while Retaining Utility.....	11
The History of Cognitive Ability Tests in the Court System .....	13
The Current Study.....	15
Criteria for Selecting Cases for Review .....	17
Coding the Cases.....	18
Results.....	19
Hypothesis 1 .....	20
Hypothesis 2.....	21
Hypothesis 3 .....	21
Hypothesis 4.....	22
Hypothesis 5.....	22
Discussion.....	23
Hypothesis 1 Findings.....	23
Findings for Remaining Hypotheses .....	27
Limitations .....	28
Conclusion.....	29
References.....	30
Appendix A: Coding Factors .....	34
Appendix B: List of Court Cases Included .....	35
Appendix C: List of Court Cases Excluded.....	37
Appendix D: Court Case Individual Coding Results .....	38

A REVIEW OF COURT CASES INVOLVING COGNITIVE ABILITY TESTING  
AND EMPLOYMENT PRACTICES: 1992-2015

Nicholas Morris

May 2016

39 Pages

Directed by: Elizabeth L. Shoenfelt, Reagan D. Brown, and Amber N. Schroeder

Department of Psychological Sciences

Western Kentucky University

This review is an extension of a study by Shoenfelt and Pedigo (2005). The purpose of this review is to help form an understanding of how the courts handle cases where an organization has used a cognitive ability test to select employees and consequently faced charges. Cognitive ability testing is the best known predictor of job performance for a wide range of jobs. However, cognitive ability testing also is known to lead organizations to select fewer members of protected groups, such as African Americans, Hispanics, and women. The cases that were reviewed were identified in the LexisNexis database. In order to review the cases, pertinent information was coded by four Industrial-Organizational Psychology graduate students then used the information as categorical data to make comparisons based on the outcome of each case and the conditions that may have led to the outcome. Findings were similar to the Shoenfelt and Pedigo (2005), which is likely due to the low number of new cases that were added to the review. Cases in which the defendant had used a validated test often ruled in favor of the defendant. However, in the six new cases that were discovered, issues such as arbitrary cutoff scores and the presentation of equally valid alternatives played a role in rulings in favor of the plaintiff even in cases with a validated test. The case claims were all race based and all involved tests that were professionally developed.

## **Introduction**

There are many considerations when determining what psychological construct will predict high performance on a particular job. The first and most important consideration for any organization is, of course, to select employees who will perform well. A well prepared organization is going to weigh the advantages and disadvantages of each of the available prediction methods. Of those available methods, there are a few that are well established as generally valid and useful predictors: personality tests, biodata, structured interviews, job samples, and integrity tests (Schmidt & Hunter, 1984). The best of these across a wide variety of job settings is general mental ability or general cognitive ability (Hough, Eaton, Dunnette, Kamp, & McCloy, 1991; Hunter & Hunter, 1984; Schmidt & Hunter, 1998).

Although it is the best, organizations that use general cognitive ability as the main predictor of job performance will likely face some issues in the form of adverse impact against African-American, Hispanic, and even female applicants (Hunter & Hunter, 1984). Adverse impact occurs when an organization's employment practices result in a disproportional number of protected group members being rejected. Because of the high potential for adverse impact when using cognitive ability as a predictor, there has been much debate over if and when it is appropriate to use it (Murphy, 2002).

In this review, all of the relevant United States appellate and district court cases from the last twenty years have been analyzed in order to find what factors influence decisions in court cases that involve an organization's use of cognitive ability testing to select employees. The purpose of doing such a review is to better understand how cognitive ability testing is being used in organizational settings and how the legal system

reacts in each unique situation. Before analyzing court decisions, it is important to cover some history of the construct of cognitive ability itself and its usefulness, the risks of using cognitive ability as a predictor of performance, and the legislation that directly affects the use of cognitive ability testing in organizations.

### **General Mental Ability/Cognitive Ability**

General cognitive ability can be summarized simply as one's ability to learn and adapt (Schmidt, 2002). As a general construct, cognitive ability encompasses a group of more specific abilities such as spatial, verbal, and numerical aptitude (Schmidt, 2002). Quite some time ago when he came up with the concept of general mental ability, Spearman (1904) realized that these more specific constructs were interrelated, and that general mental ability is central to all human affairs. Since then, there has been extensive research on the construct of mental/cognitive ability. For example, there is some support that there is a very strong correlation (ranging from .5 to .8) between different types of tests of cognitive ability (Carroll, 1993, as cited in Hunt, 1996). Because there is some degree of relatedness between all of these specific cognitive ability measures, a test of general cognitive ability would likely predict performance about as well as a more specific measure, such as spatial ability. The universally high quality of workers selected through cognitive ability testing is logically consistent with the idea that those who score high on such a test are generally good at navigating human affairs (Schmidt, 2002).

### **Cognitive Ability as a Predictor of Job Performance**

Through an extensive meta-analysis of standalone predictors, Hunter and Hunter (1984) laid an impressive foundation of knowledge for a considerable range of available predictors. For cognitive ability testing, they found that not only did it predict success in

training for all job families with a validity estimate of .55, but that it had a validity coefficient of .61 with job performance in the most complex job families and still maintained a coefficient of .27 in the least cognitively complex jobs (Hunter & Hunter, 1984). When comparing predictors, Hunter and Hunter found that combining cognitive ability and psychomotor ability tests across all of the job families they examined provided validity coefficients ranging from .43 to .62, which represent high levels of predictive validity and, as a result, utility.

Schmidt and Hunter (1998) performed a follow up study in which they analyzed combinations of cognitive ability testing and other predictors and used standardized multiple regression to determine incremental validity of the additional predictors. The predictors they found to be the most effective when combined with cognitive ability were integrity tests, work sample tests, structured interviews, and conscientiousness tests (Schmidt & Hunter, 1998). Apart from job complexity, not many moderators of the relationship between cognitive ability and job performance are known to exist (Gutemberg, Arvey, Osburn, & Jeanneret, 1983; Murphy, 2002). This further supports the notion that cognitive ability is a strong predictor, even in various situations that have different factors determining how well employees perform. As a side note, Hunter, Schmidt, and Le (2006) found that, due to suboptimal corrections for range restriction in the then current meta-analysis process, meta-analytic studies of cognitive ability results may actually have underestimated the validity coefficients for cognitive ability testing by as much as 25%. This proposed deficiency in the meta-analysis process could mean that the true relationship between cognitive ability and job performance might be even stronger than the Hunter and Hunter (1984) analysis suggested.

## **Disagreement of General Cognitive Ability as a Predictor of Performance**

Although there is much support for the predictive value of cognitive ability, there is some disagreement that cognitive ability tests are actually the best predictor of job performance. Sternberg and Wagner (1993) believed that the utility of general cognitive ability tests was more glorified than it should be. Particularly, they stated that it is more important to measure what they called tacit knowledge, a more specific and practical kind of intelligence that involves the use of knowledge gained while on the job. According to Sternberg and Wagner, tacit knowledge is correlated from .30 to .50 with different measures of job performance across a variety of jobs (1993).

McClelland (1993) likewise disagreed with the glorification of cognitive ability testing. He felt that there may be some third variable, such as race, background, or education that is more heavily influencing job success than cognitive abilities. McClelland thought that if that third variable was to be added to the analysis as a control, there would be less predictive value in cognitive ability when predicting job performance. However, for the purposes of predicting one's potential for high supervisory ratings/job performance, knowing that there is a relationship between the two, measured as they are, is, according to the Uniform Guidelines on Employee Selection Procedures, enough to serve as evidence of the validity of cognitive ability tests for personnel decision making (EEOC, 1978).

Both McClelland (1993) and Sternberg and Wagner (1993) presented interesting arguments against the scientific accuracy of some of the claims made about cognitive ability testing, but they do not detract from the overall value that can be derived from the use of cognitive ability tests in selection. Even though there may be more to the

relationship between performance and different factors surrounding intelligence, general mental ability or general intelligence still has overwhelmingly strong support as a predictor in applied settings (Hough, Eaton, Dunnette, Kamp, & McCloy, 1991; Hunter & Hunter, 1984; Schmidt & Hunter, 1998).

### **The Risk of Using Cognitive Ability Tests**

The major problem with cognitive ability testing is that there is evidence that there are large significant differences in racial (Hunter & Hunter, 1984) and female subgroups (Ployhart & Holtz, 2008) on certain measures of cognitive ability when scores are compared to those of the Caucasian male subgroup. African Americans and Hispanics suffer the most adverse impact from these differences. African Americans tend to score about one standard deviation lower than Caucasians score. The second most affected group is Hispanics. Hispanic Americans typically score slightly higher than African Americans (Murphy, 2002; Outtz, 2002). Even women have a chance of being adversely impacted by cognitive ability tests. Women tend to score lower on ability tests that focus more on spatial intelligence or, to a lesser extent, quantitative ability (Hough, Oswald, & Ployhart, 2001). More often than not, African Americans are the most adversely impacted group, consistent with the finding that their scores are typically the lowest on general cognitive ability tests (Hunter & Hunter, 1984).

Because these differences exist, it is hard to use cognitive ability testing without resulting adverse impact against minorities. There is not only risk of legal trouble for discriminating, an organization also must consider how it is viewed by the world. If there is a concern for social equity within an organization, it would be best to pair cognitive ability testing with another predictor that causes less adverse impact or to just avoid

cognitive ability testing altogether. Previously, it was suggested to select top-down on cognitive ability within racial subgroups as a way of taking affirmative action (Schmidt & Hunter, 1998). However, doing so is now illegal since the passage of the Civil Rights Act of 1991. The Civil Rights Acts will be discussed later in more detail.

### **Should an Organization Use Cognitive Ability Testing?**

As one would expect, the decision whether or not to use the best predictor of performance that has an almost certainty of race-based adverse impact has been the source of much debate. Murphy (2002) divided the debate into two camps: one that favors organizational efficiency, performance, productivity, or profit and the other that favors group parity, non-discrimination, equity, or social goals. Murphy also stated that, at its core, the separation is defined by a difference in values (2002). If an organization chooses to rely on cognitive ability testing for its selection procedures without any steps to prevent adverse impact, there will almost certainly be some adverse impact regardless of whether or not the organization has ill intentions toward those who are not Caucasian and male. Murphy (2002) contended that this discrimination is legally defensible when an organization has proof of validity; that is, that their methods involve measuring what will be important to performance on the job. However, Murphy felt that making this decision is still a commitment to live with the social consequences of the test. Another important ethical and legal note is that if there is an equivalently valid test that could be used as a predictor that does not have adverse impact, the organization in question should use it (Outz, 2002). Also, choosing social equity over organizational efficiency is a good way to emphasize an organization's commitment to its workers and their rights (Murphy, 2002).

The race-based differences in performance on cognitive ability tests are not due to test bias alone (Gottfredson, 2000). They are due to real differences in the ability of different races. One important difference is in literacy. Gottfredson noted that only 1 of 30 African-Americans function at level 4 or above on the U.S. Department of Education's National Adult Literacy Survey, whereas 1 of 4 Whites do. Most African-Americans fall at levels 1 and 2, which means they are unlikely to be able to perform fully as a citizen or in the global economy. According to Gottfredson, 40% of Caucasian adults fall in this range while 75% percent of African Americans do. In other words, this study found that there is a 35% difference in the number of African Americans that cannot read well enough to perform fully as American citizens and the number of Caucasians who can. As Gottfredson stated, that is a big difference in literacy. This difference in literacy means that many African Americans who are asked to take a general cognitive ability test that involves any amount of reading could have more difficulty completing it than would a Caucasian applicant. If African Americans cannot understand the test questions, they likely cannot answer in a way that would earn a high score. Ployhart and Holtz (2008) suggested that lowering the reading ability requirements of the tests used would likely decrease the size of subgroup differences on tests of cognitive ability. That suggestion is likely based on findings about differences in reading ability such as those that Gottfredson (2000) discussed. However, although it is important to keep this difference in literacy in mind, it is also important to note that differences in literacy alone do not settle the debate about what the underlying cause of the subgroup differences of minorities on cognitive ability tests.

As mentioned briefly before, there are important consequences involving public perception of social equity that need to be considered when an organization chooses cognitive ability as a predictor. However, the organization is likely to be more focused on the legal implications of adverse impact. In some cases, organizations are made to pay damages, back pay, or even hire those who have been adversely impacted by the employment practices. The amount paid for these sentences can be millions or even tens of millions of dollars (Williams, Shaffer, & Ellis, 2013).

### **Important Court Cases and Legislation**

Given the long history and use of general cognitive ability tests (Hunter & Hunter, 1984), there has been plenty of time for the courts to uncover and resolve disputes over discrimination and develop societal guidelines to maintain professional ethics and avoid disparate impact. If an organization decides to use general cognitive ability tests and take the risks involved, then they should start by learning about the major laws and precedents that now dictate how the courts handle any issues that would arise.

**Civil Right Acts of 1964 and 1972.** Title VII of the Civil Rights Acts of 1964, as amended in 1972, was intended to prevent organizational discrimination based on an employee's sex, color, race, religion, or national origin. It also established that the initial burden was on the plaintiff (i.e., the employees being affected) to provide evidence. As is evident in the cases that followed, the Civil Rights Act of 1964, as amended in 1972, was not enough to stop organizations from using selection practices that caused adverse impact.

***Griggs v. Duke Power Company (1971).*** Perhaps the first turning point of a case to impact the law surrounding cognitive ability testing and racial differences was the

Supreme Court case, *Griggs v. Duke Power Company* (1971), where a power company had decided to divide its workers into two groups: one with low pay and hard labor duties and one with better pay and operational duties. The criteria to be admitted into the higher paying group required that the applicant had a high school diploma and was able to pass several cognitive tests. This requirement resulted in 94% of African American employees who applied for the higher pay group being denied. Meanwhile, Caucasians were admitted into this pay group without passing the cognitive tests but with a high school diploma; 42% of the Caucasian applicant pool was denied (*Griggs v. Duke Power Company*, 1971). In the end, it was decided that the needs of the job did not necessitate either of the selection criteria used. This case established that tests had to be reasonably job related. This case is part of the reason why courts now look at validity evidence, as it is a measure of job relatedness.

**Uniform Guidelines on Employee Selection Procedures.** After many disputes, the Equal Employment Opportunity Commission created a set of guidelines known as the Uniform Guidelines on Employee Selection Procedures (Equal Employment Opportunity Commission [EEOC], 1978). These guidelines are not meant to be interpreted as law but a guidance for good practice when implementing selection tools because they are given great deference by the courts. The Uniform Guidelines (EEOC, 1978) outline what it takes to properly validate a test and what indicates the occurrence of adverse impact. According to the Guidelines, adverse impact occurs when the selection ratio of a protected subgroup of applicants falls below 80% of the selection ratio for the group with the highest selection ratio. Also, the types of validity studies that are recommended are criterion, content, or construct related validity studies (EEOC, 1978). These are all

studies that involve some form of objective support for the relationship between performance on required job duties and whatever predictor the organization may be using. As a nod to psychology practitioners and selection predictor test developers, the Guidelines include a brief statement that the guidelines themselves are consistent with good professional practice in these fields.

**The Civil Rights Act of 1991.** An important step in this history is the passing of the Civil Rights Act of 1991, an addendum to the original Civil Rights Acts of 1964 and 1972, which came about as the result of several discrepant rulings in discrimination cases.

The first important case is *Patterson v. McLean Credit Union* (1988). In this case, several African American employees were caused emotional damages but were unable to receive any monetary settlement for their discriminatory treatment because the employer had not denied them the same rights as white citizens in terms of contracts for work. Of course, there was little justice in this case as these employees were left with evidence of discrimination and nothing rectified it. After the Civil Rights Act of 1991 passed, terms were changed in the original law used against the African Americans in this case to better represent the rights that were to be given to African American workers in terms of contracts. The Civil Rights Act of 1991 read that they were entitled the same right to “the making, performance, modification, and termination of contracts, and the enjoyment of all benefits, privileges, terms, and conditions of the contractual relationship” as Caucasian citizens.

The next highly relevant cases were *Wards Cove Packing Co. v. Antonio* (1989) and *Price Waterhouse v. Hopkins* (1989). *The Wards Cove Packing Co. v. Antonio* case

made it clear that the burden of providing proof that an organization's personnel practice would have disparate impact should not fall on the plaintiff, but the organization itself should provide evidence that their personnel practice is valid and consistent with business necessity. The Civil Rights Act of 1991 recognized that it was far too difficult for a plaintiff to win a case against an organization that legitimately had discriminated against them. To balance out the fairness of this new change, Congress added that an imbalance in the organization's workforce in terms of race, sex, or other protected groups was not enough to make a prima facie case of disparate impact. The *Price Waterhouse v. Hopkins* (1989) case made Congress realize that an organization should not be treated leniently for disparate treatment by stating that they would have made the same decision regardless of whether the affected individual was in a protected group or not. The subsequent change was that organizations would be able to avoid back pay and reinstatement but not attorney and court fees if they could show they would have made the same decision regardless of protected group status.

The main impact the Civil Rights Act of 1991 had on organizational personnel procedures is the reinforcement of the need to validate predictors either through validity generalization or a validity study specific to the organization potentially in question. With proper statistical validation comes objective evidence of job relatedness, which is something that any organization using cognitive ability testing is going to need in preparation for any potential legal challenges.

### **Reducing Risk of Adverse Impact while Retaining Utility**

There are several proposed strategies for limiting subgroup differences. Some involve using methods that include a more comprehensive look at an employee's ability

such as interviews or assessment centers, choosing applicants from within bands, and minimizing the verbal ability requirements of the tests used (Ployhart & Holtz, 2008). However, these are likely to somewhat decrease validity. The only strategy that Ployhart and Holtz found that did not decrease validity was to assess applicants on all of the knowledge, skills, and abilities required to perform the job in a battery of tests (2008). However, that may prove to be a difficult measure to develop and administer in consideration of the resources available to most organizations and how much it would cost to develop each test.

One has to be careful when considering alternative measurements of general cognitive ability that do not have the same disparate impact on subgroups. One study attempted to use logical reasoning tests in place of established cognitive ability tests (Soete, Lievens, & Druart, 2013) to explore an option that would likely have less adverse impact. In eliminating the impact, it is possible that the construct that is measured might be fundamentally different from cognitive ability. To clarify, individuals that may have a low score on a test of cognitive ability could have a much higher score on an alternative test, one that attempts to measure the same overarching construct that cognitive ability tests do. Even though applicants scored high on the alternative, their performance levels may still be as low as they were predicted to be with the cognitive ability test. There is potential for this to happen because there are fundamental differences in the construct of cognitive ability between racial groups presented in several studies (Gottfredson, 2000; Schmidt, 2002). If those differences in test scores no longer exist, it is possible that general cognitive ability is no longer being measured. This was demonstrated in the Soete et al. study (2013).

Contextualization is another strategy where circumstantial and situational information is provided with the questions on the cognitive ability test (Soete et al., 2013). This format for the questions makes them easier to conceptualize for applicants and has the potential to frame the test as something that is more easily perceived as job related than a generic test with vague questions. Another similar strategy is the use of simulations such as a sample of the work that will be done on the job or a situational judgment test where applicants are asked to make a judgment about what to do to resolve some work problem scenario. Combining these two with cognitive ability also has been found to be effective for validity and to decrease the size of subgroup differences in situations where the work sample and situational judgment test are created in a way that does not also cause adverse impact (Soete et al., 2013). An example of what could cause the combination of cognitive ability tests and situational judgment tests to still result in adverse impact would be using examples with a high verbal requirement (Gottfredson, 2000; Ployhart & Holtz, 2008).

### **The History of Cognitive Ability Tests in the Court System**

It is vitally important to keep legal defensibility in mind when using cognitive ability testing. As mentioned before, the costs of losing a court case can be crippling for some organizations. The literature on cognitive ability testing provides a few good ways to increase the legal defensibility of an organization's selection process.

Terpstra, Mohamed, and Kethley (1999) examined federal court cases involving cognitive ability tests and took a closer look at several performance predictors' tradeoff between validity/utility and their legal risk. In their examination of cognitive ability tests, they found that 28 of the court cases from 1978 to 1997 were challenging the use of

cognitive ability tests and 14 of those were more specific mental ability test cases; four involved more general intelligence or aptitude test cases (Terpstra et al., 1999).

Perceptions of job relatedness seem to be important for determining whether or not a person will decide to pursue discrimination charges, given that the more specific tests of mental ability were challenged far more than were the general tests. With that in mind, it may be important to keep using more general tests of cognitive ability and it probably would help to include some job related information in the test itself by using a process such as contextualization (Soete et al., 2013).

As mentioned in the Uniform Guidelines (EEOC, 1978), validity evidence of some form is required to support the use of whichever predictor an organization may be defending in court. Validity is an important factor in determining how well a selection tool will hold up in court as it provides objective evidence concerning how the selection tool was determined to be related to job duties. In a review of court cases dealing primarily with cognitive ability tests from 1991 to 2004, Shoenfelt and Pedigo (2005) found that the court ruling was typically in favor of the organization when the test used was standardized and validated. In the few cases where there was a valid test but the court ruled in favor of the plaintiff, it was due to the way the organization set their cutoff scores for selection. Therefore, it is highly important that every decision that is made concerning a selection process be based on objective job data (Shoenfelt & Pedigo, 2005). They also suggested that if an organization does not have the resources to have an in-house practitioner validate their selection test, they should consider hiring one to avoid the legal damages that might come from using a test with no objective connection to job

performance. The only case in the Shoenfelt and Pedigo study (2005) without a validation study ruled in favor of the plaintiff and the organization was made to pay damages.

When considering options other than cognitive ability tests, it is important to consider which among those other predictors also might be legally risky. Other selection tools are challenged in court almost as often as, or more often than, cognitive ability testing. Terpstra et al. (1999) found that in court cases sampled from 1978 through 1997, there were 91 challenges to unstructured interview processes, 9 challenges to structure interviews, 28 to cognitive ability testing, 22 to physical ability tests, and 7 to work sample tests. In their study, unstructured interviews were the least likely to find support in court with the defendant succeeding at a rate of 59%. For comparison, cognitive ability tests overcame 67% of challenges and work samples overcame 86% of challenges (Terpstra et al., 1999). Some of the other predictors identified by Schmidt and Hunter (1998) as having considerably high predictive validity were work sample tests (.51), integrity tests (.41), structured interviews (.51), and tests of conscientiousness (.31). Any of those could prove useful in predicting worker performance and they likely do not face as many challenges in court as does cognitive ability.

### **The Current Study**

Although there are less risky options, cognitive ability testing is still desirable as it is a universally applicable predictor of performance (Hunter & Hunter, 1984). When cognitive ability is used and the organization is faced with a person or group that has been adversely impacted, the organization may have to appear in court if they do not settle the case. The purpose of the present study is to examine existing court cases to

determine what patterns there are in the court decisions regarding the use of cognitive ability testing as a selection tool.

As established by *Griggs v. Duke Power* (1971) and the Uniform Guidelines (EEOC, 1978), the courts have recognized that when a selection test has been found to cause adverse impact, the organization must have validity evidence. Therefore, it is expected that:

*Hypothesis 1:* The proportion of court cases that rule in favor of the defendant when a properly validated test is used is expected to be higher than the proportion of court cases that rule in favor of the defendant when there is not a properly validated test.

According to Hunter and Hunter (1984), minority groups are very likely to be negatively impacted by intelligence testing. Also, because race is a protected group, it is likely that African-Americans and Hispanics will be plaintiffs in appeals and district courts. Therefore, it is expected that:

*Hypothesis 2:* The proportion of cases with plaintiffs who are minority group members will be greater than the proportion of cases with plaintiffs who are not members of minority groups.

Although women are considered to be negatively affected by cognitive ability tests, it is to a much lesser degree than African Americans or Hispanics (Ployhart & Holtz, 2008). Therefore, it is expected that:

*Hypothesis 3:* The proportion of race-based claims will be greater than the proportion of gender-based claims.

In the Uniform Guidelines on Employee Selection Procedures (EEOC, 1978), having validated tests is important. The Uniform Guidelines also state that tests need to be professionally developed. A test can be “validated” but not developed by someone who has the proper abilities to do so. Therefore, it is expected that:

*Hypothesis 4:* The proportion of cases ruled in favor of the defendant with a professionally developed test will be higher than the proportion of cases ruled in favor of the defendant without a professionally developed test.

Terpstra et al. (1999) found that cases with jury trials typically involved much larger sums of money being paid in damages for cases that the defendant lost. This implies that the jury sympathizes to some degree with the plaintiff. Therefore, it is expected that:

*Hypothesis 5:* The proportion of cases where the court ruled in favor of the plaintiff and there is a jury present will be greater than the proportion of cases where the court ruled in favor off the plaintiff without a jury trial.

## **Method**

### **Criteria for Selecting Cases for Review**

As an extension of the Shoenfelt and Pedigo (2005) review, this study used a similar method to select cases for review. The cases reviewed are from the appellate and district court levels. To find relevant cases, the researcher utilized Lexis-Nexis, a search engine for legal documents. Shoenfelt and Pedigo (2005) searched the years 1992 to 2004. They searched only as far back as 1992 because of the changes in court decisions subsequent to the 1991 Civil Right Act. This review covers the cases that were decided between January 1992 and December of 2015 and includes most of the cases that were

reviewed in the Shoenfelt and Pedigo (2005) study. The following words, either alone or in combination, were used in Lexis-Nexis in order to find relevant cases: employment testing, test validation, content validity, criterion validity, construct validity, cognitive testing, cognitive ability testing, intelligence testing, and selection.

The search returned very few new cases; only six new cases were found. Two of the cases from the previous review appeared in court again (*USA v. The State of Delaware*, 2004; *Banos v. City of Chicago*, 2004), but the ruling on the test itself did not change in either. The cases that Shoenfelt and Pedigo (2005) determined should be left out of their analysis were also left out of the analysis in this review. In addition, two more cases from the previous review were excluded as it was determined they were not directly related to the cognitive ability test itself but to some other test or factor. In *Brunet v. City of Columbus* (1995) the issue was the physical ability test being weighted too heavily in the scoring process. Although there was a cognitive ability test present, it was not of issue. In *Jordan v. City of New London* (1999), the nonminority plaintiff complained that he was discriminated against for scoring too high on the test. This case was determined to be irrelevant because it did not involve a minority group or the typical issues that cognitive ability testing can cause. In the end, the total number of cases in this review was 24, of which 9 were appellate court cases and the other 15 were from the district courts (see Appendix D). Cases from Shoenfelt and Pedigo excluded from the current study may be found in Appendix C.

### **Coding the Cases**

This review used the same coding method used in Shoenfelt and Pedigo (2005) that was adapted from the Werner and Bolino (1997) study that reviewed cases

concerning performance appraisal and the Terpstra et al. (1999) review of cognitive ability test cases. In addition to coding factors that were used before, this study included the factor of cutoff scores as it was determined to be a potentially deciding factor in whether the court ruled in favor of the defendant in the Shoenfelt and Pedigo (2005) study. The factors that were coded can be seen in Appendix A.

There were four raters, all of whom were industrial-organizational psychology graduate students. There were three independent raters for each of the six new cases. The cases from the previous review were recoded. Only one rater was deemed necessary for the previously reviewed cases to check that the coding in the last review was accurate and to code the additional factors added in this review. An evaluation of rater agreement was conducted.

## **Results**

The coding results for all of the individual cases are included in Appendix D. There was virtually no disagreement among the raters in coding the cases. In a few cases there was some confusion about whether a test with multiple components was considered to be multiple tests or a single test. In the end, that factor was coded as multiple tests if there were several dimensions being measured in one test battery. In coding the cases from Shoenfelt and Pedigo (2005), the current raters had no disagreement with the raters from that review. There were a few instances of rater disagreement on the six new cases; these disagreements were resolved after discussing the rationale behind the coding factor and discovering a misunderstanding about how to interpret the code.

In order to test Hypotheses 1, 4, and 5, a z-test of the proportions was used to analyze if there were significant difference in the proportions of cases as hypothesized.

The formula for this test is provided below<sup>1</sup>:

$$z = \frac{\frac{r1}{n1} - \frac{r2}{n2}}{\sqrt{\left(\frac{r1 + r2}{n1 + n2}\right) \left[1 - \left(\frac{r1 + r2}{n1 + n2}\right)\right] \left(\frac{1}{n1} + \frac{1}{n2}\right)}}$$

Where, in this review:

*r1* is the number of cases that meet the relevant criteria

*n1* is the total number of cases in relevant comparison group 1

*r2* is the number of cases that meet the relevant criteria

*n2* is the total number of cases in relevant comparison group 2

To clarify how Hypotheses 1, 4, and 5 were tested, each component of the analysis for that hypothesis is restated in the equation below the hypothesis. Hypotheses 2 and 3 were analyzed using a z-test to determine if the observed proportion of cases was significantly different from a hypothetical distribution where the cases were evenly split on the relevant criteria.

### **Hypothesis 1**

Hypothesis 1 was stated as: The proportion of court cases that rule in favor of the defendant when a properly validated test is used is expected to be higher than the proportion of court cases that rule in favor of the defendant when there is not a properly validated test.

In order to test Hypothesis 1, the proportions compared were as follows:

---

<sup>1</sup> z-tests were computed using an online calculator after it was determined that the online calculator provided the same results as the z-test formula calculated by hand.  
<http://www.socscistatistics.com/tests/ztest/Default2.aspx>

$$\frac{\text{Cases validated and in favor of defendant } (r1)}{\text{Total cases validated } (n1)}$$

vs

$$\frac{\text{Cases not validated and in favor of defendant } (r2)}{\text{Total cases not validated } (n2)}$$

The z-test for differences in proportions indicated there was a significant difference between the number of cases in favor of the defendant when there was a validated test being used (16/22) and cases in favor of the defendant when there was not a validated test being used (0/2;  $z = 2.19, p < .05$ ). There were significantly more cases ruled in favor of the defendant when the test being challenged was validated than there were when the test was not validated, thus supporting Hypothesis 1.

### **Hypothesis 2**

Hypothesis 2 stated: The proportion of cases with plaintiffs who are minority group members will be greater than the proportion of cases with plaintiffs who are not members of minority groups.

In order to test Hypothesis 2, a z-test was performed. The proportion of cases with plaintiffs who were of a minority race (22/24) was significantly different than if minorities and non-minority plaintiffs were equally represented in the court cases observed (12/24;  $z = 4.279, p < .01$ ). There were significantly more cases with minority plaintiffs than nonminority plaintiffs, thus supporting Hypothesis 2.

### **Hypothesis 3**

Hypothesis 3 stated: The proportion of race-based claims will be greater than the proportion of gender-based claims.

There were no gender-based cases found. All 24 cases in this review were race-based cases, and a z-test confirmed that this finding is significant ( $z = 5.10, p < .01$ ).

Therefore, Hypothesis 3 is supported.

#### **Hypothesis 4**

Hypothesis 4 stated: The proportion of cases ruled in favor of the defendant with a professionally developed test will be higher than the proportion of cases ruled in favor of the defendant without a professionally developed test.

$$\frac{\text{Cases for defendant with prof dev tests } (r1)}{\text{All cases with prof dev tests } (n1)}$$

vs

$$\frac{\text{Cases for defendant without prof dev test } (r2)}{\text{All cases without prof dev tests } (n2)}$$

All 24 cases involved professionally developed tests. Of these cases, 16 of the 24 were decided for the defendant. The z-test indicated that the proportion of cases in favor of the defendant was significantly different from the proportion of cases where the defendant won and they had not used a professionally developed test (0) ( $z = 4.50, p < .01$ ). Thus, Hypothesis 4 was supported.

#### **Hypothesis 5**

Hypothesis 5 stated: The proportion of cases where the court ruled in favor of the plaintiff and there is a jury present will be greater than the proportion of cases where the court ruled in favor off the plaintiff without a jury trial.

$$\frac{\text{For plaintiff with jury trial } (r1)}{\text{Cases with jury trials } (n1)}$$

vs

$$\frac{\textit{For plaintiff without jury (r2)}}{\textit{Cases without jury trials (n2)}}$$

Unfortunately, there was only one court case that involved a jury trial. Therefore, it would be misleading to do any analysis. Given that there was only one isolated case with a jury trial and nothing with which to compare it, there is insufficient information to test Hypothesis 5.

## **Discussion**

### **Hypothesis 1 Findings**

Hypothesis 1, which stated that the proportion of court cases that rule in favor of the defendant when a properly validated test is used is expected to be higher than the proportion of court cases that rule in favor of the defendant when there is not a properly validated test, was supported. It is always advisable to perform a job analysis for a job associated with employment decisions and to obtain validity evidence for the testing process being used to make those decisions. Support for validated tests is beneficial for industrial-organizational psychology practitioners, as this finding is consistent with the Uniform Guidelines (EEOC, 1978) and that the courts give great deference to the Guidelines. The number of cases decided in favor of the plaintiff (6/22) in this comparison is particularly interesting. The tests being defended were validated in all six of these cases, and yet the plaintiff won. The most prevalent reason for the plaintiff's victory was that the cutoff scores were not supported by the job analysis data or any other professional standard. The other reason, found only in *Bazile v. City of Houston* (2012), was that an equally valid alternative existed to the cognitive test used. It is worth noting that the Bazile case was the only one in which the plaintiff managed to provide a more valid alternative and to win the case. Again, this argument by the plaintiff, that is, a

viable alternative with equivalent validity and less adverse impact, is consistent with the Guidelines (1978). In the other five cases with validated tests that were found for the plaintiff, the reason the defendant lost the adverse impact case was having set a cutoff score that did not align with the job analysis data. Because this was such a prevalent finding in the recent cases, the next section addresses the particular issues in each case.

### **Cutoff Scores and Other Influences on the Judgment of Validated Tests.**

In the original case from Shoenfelt and Pedigo (2005) that inspired the new coding factor for cutoff scores in the current review, *Green v. Town of Hamden* (1999), there was a ruling for a preliminary injunction for the defendant's selection process because their cutoff score did not serve any legitimate business necessity. With a set cutoff of 75%, only one minority applicant achieved a passing score. The score was then moved to 60% when this problem became apparent. However, no professionally accepted standard or methodology was used in determining these percentage cutoffs. The professional who developed the test did not make any recommendations regarding a specific cutoff score when the test was validated because there were differences among the municipalities that might use the test. The professional who developed the test did, however, provide some factors to consider when determining a cutoff score in these different situations: the number of openings, performance on the test, impact, and diversity (*Green v. Town of Hamden*, 1999).

The cutoff score in *Lewis v. City of Chicago* (2005) was the primary reason the defendants lost with a validated test. In the job analysis conducted by the industrial-organizational psychologist, it was determined that the city should set the cutoff score by counting down from the top scores in increments of 13 because he had determined that

scores within 13 points of one another were not significantly different. The cutoff score of 89 set in this case was not within that band, therefore it was determined to be arbitrarily set. The test's reliability also was called into question. This case cites *Lanning v. SEPTA* (1999), a court precedent concerning cutoff scores in discrimination where the judge ruled that a cutoff score on an entry-level employment examination must be shown to measure the minimum qualifications necessary for successful performance of the job in question. Also, *Lewis vs. Chicago* (2005) was overturned in 2008 because of the length of time that had passed since the complaints were filed (Dunleavy & Gutman, 2009). The judges in the appealed case ruled that the timely filing period for adverse impact cases is when scores were made known to the test takers, not when the decisions were made based on the inferences from the scores (Dunleavy & Gutman, 2009). The timeliness of complaints is not the focus of this review; because the case was initially decided on the job relatedness of the test, our analysis is still accurate. This timeliness trend should be noted. However, in the end, the best advice for an organization is to spend time and effort to create reliable and validated selection procedures up front that also attempt to minimize adverse impact (Dunleavy & Gutman, 2009).

The *Bazile v. City of Houston* (2012), a case found for the plaintiff, is of interest for a reason other than the cutoff score. Adverse impact was demonstrated using the 4/5ths rule, but no statistical tests were significant. Although the more rigorous analyses suggested the 4/5ths violation occurred by chance, the court found adverse impact (Dunleavy, 2012). In the end, the court ruled that the equally valid assessment center and situational judgment test be used, but not given any greater weight than that for the multiple-choice test (*Bazile v. City of Houston*, 2012).

In *Smith v. City of Boston* (2015), the city had set a cutoff score of 70 for no identified reason. However, this was not the main problem in the court's opinion. The city of Boston had used this test for a ranking system, and the court deemed that the validity evidence did not support rank ordering (i.e., that an applicant with better scores on that test would perform better on the job). Thus the rank ordering of applicants based on their test was unlawful because it resulted in adverse impact (*Smith v. City of Boston*, 2015). Note that the coding for Smith (see Appendix D) may make it appear that the reason the plaintiff won was the cutoff score issue; however, it was actually the rank-ordering of applicants that led the court to rule in favor of the plaintiff.

In *Local Union v. Mississippi Power and Light* (2004), the cutoff score was based on data, but was found to cause adverse impact. The court ruled that Mississippi Power and Light could no longer to use the test until they had computed a proper cutoff score that met the businesses needs but also caused less adverse impact (*Local Union v. Mississippi Power and Light*, 2004)

In *United States of America v. City of New York* (2010), the court determined that there was no relation between the cutoff score and the job analysis data. It was ruled that there was intentional discrimination across time, otherwise known as pattern or practice. In this case, despite having a validated test, City of New York lost because of the pattern or practice evidence presented and a rank-ordered pass-fail test with a cutoff score that misrepresented the job data (*United States of America v. City of New York*, 2010).

Together, these cases with cutoff score issues provide further support for the statement by Dunleavy and Gutman (2009) regarding *Lewis v. City of Chicago* (2005).

That is, subtest weighting, cutoff scores, and banding strategies should all be demonstrably empirically based.

### **Findings for Remaining Hypotheses**

Hypothesis 2 was supported. There were significantly more cases with minority plaintiffs than nonminority plaintiffs. There were few reverse discrimination cases and no gender-related cases. This finding was expected because women typically do not differ from men in cognitive ability. Race-based differences in cognitive ability, particularly for minorities such as African-Americans and Hispanics, are common (Ployhart & Holtz, 2008).

There were no gender-related cases identified in the past 20 years. This is substantive information on its own. Hypothesis 3 was supported, as all 24 cases were race based. There was a gender-based case (*Brunet v. City of Columbus*, 1995) in the previous review (Shoenfelt & Pedigo, 2005), but after looking more closely at the content of the case, it became clear that the focus was the physical ability test and not the cognitive ability test. Thus, *Brunet v. City of Columbus* was removed from the analysis of the current study.

Hypothesis 4 was supported, as the defendant won in 16 of the 24 cases where the test was professionally developed. There were no cases in which the test was not professionally developed. Thus, it can be reasoned that it is favorable for the organization to use a professionally developed test. However, as illustrated by the five cases that turned on the cutoff score, organizational decisions on how to implement a professionally developed test can lead to litigation.

Only one case (*Zottola v. City of Oakland*, 2002) involved a jury trial. Thus, Hypothesis 5 could not be tested in the current study, nor was it tested in Shoenfelt and Pedigo (2005). The *Zottola v. City of Oakland* (2002) case is unique among the other cases found; it was the only one that involved adverse impact as well as disparate treatment claims. The Civil Rights Act of 1991 indicates that only disparate treatment cases may be heard by a jury, and most of the cases gathered appeared to be adverse impact claims.

### **Limitations**

The major limitation of this review is that there were surprisingly few cases discovered. Only six court cases involving cognitive ability tests were found since 2004, an exceptionally low number of cases to have occurred across a 10-year span. There were 18 cases in the ten years covered in the Shoenfelt and Pedigo (2005) review. It is possible, but not probable, that some relevant cases were missed. The search for the current study was rigorous. However, if a search engine other than LexisNexis were used, additional cases might be found.

Another reason for the small number of cases could be that more cases are settling out of court than before. Organizations may prefer to keep cases to a minimum for the preservation of the organization's reputation. In future reviews, it would be wise to use a more overarching or encompassing process for finding cases involving more people and more search engines such as Westlaw or Google. It also may be informative to include a qualitative analysis of the cases while still addressing the quantitative aspect for drawing conclusions.

For any future review of court cases concerning cognitive ability tests, it would be wise to search for terms such as “written exam” or “multiple choice.” These terms may yield more results than the terms used in this review. It also might be beneficial to include extra coding factors. One such factor could be the type of validity (content, criterion, generalized, or construct) used by the defendant. Another factor could be whether the plaintiff successfully presented an equally valid alternative test with less potential for adverse impact. In addition, adding an additional coding factor addressing the type of discrimination (i.e., adverse impact, pattern or practice, or disparate treatment) associated with the selection practice or test might be of interest. Additional information potentially could increase the accuracy of any future reviews.

## **Conclusion**

This review confirms what are considered to be best practices in industrial-organizational psychology and that are consistent with the law. The results of the study indicate defendants with validated tests are likely to be successful in court unless they have arbitrarily set cutoff scores or neglected to search for equally valid alternatives with less adverse impact. The current practice of using job analysis as the basis for professionally developing and validating tests appears to be very prevalent, at least in the civil service organizations involved in the reviewed cases. Validated tests based on job analysis are likely to be upheld in court, consistent with practices described in the Uniform Guidelines (EEOC, 1978) and other professional standards. It is advantageous for organizations to make data-based decisions and to use empirically-based methods for determining sensitive selection parameters such as cutoff scores and bands.

## References

- Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241 (1964).
- Civil Rights Act of 1971, Pub. L. No. 42-22, 17 Stat. 13 (1971).
- Civil Rights Act of 1991, Pub. L. No. 102-166, 150 Stat. 1071 (1991).
- Dunleavy, E. (2012). A deeper dive into the adverse impact analyses in *Bazile v. City of Houston* (2012). *DCI Consulting Blog*. <http://dciconsult.com/a-deeper-dive-into-the-adverse-impact-analyses-in-bazile-v-city-of-houston-2012/>
- Dunleavy, E. & Gutman, A. (2009). *Lewis v. City of Chicago: What is a timely claim of disparate impact? On the Legal Front. The Industrial-Organizational Psychologist*. <http://www.siop.org/tip/jan10/08gutman.aspx>
- Equal Employment Opportunity Commission. (1978). *Uniform guidelines on employee selection procedures*. 29 C.F.R. 1605.
- Gottfredson, L. S. (2000). Skills gaps, not tests, make racial proportionality impossible. *Psychology, Public Policy, and Law*, 6, 129-143. doi:10.1037/1076-8971.6.1.129
- Griggs v. Duke Power Company* (1971) 401 U.S. 424
- Gutentberg, R. L., Arvey, R. D., Osburn, H. G., & Jeanneret, P. R. (1983). Moderating effects of decision making/information processing job dimensions on test validities. *Journal of Applied Psychology*, 68, 602–608. doi:10.1037/0021-9010.68.4.602
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1991). Criterion-related validities of personality constructs and the effects of response distortion on those validities. *Journal of Applied Psychology*, 75, 581–595. doi:10.1037/0021-9010.75.5.581

- Hough, L. M., Oswald, F. L., Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9, 152-194. doi: 1.11111468-2389.00171
- Hunt, E. (1996). When should we shoot the messenger? Issues involving cognitive testing, public policy, and the law? *Psychology, Public Policy, And Law*, 2, 486-505. doi:10.1037/1076-8971.2.3-4.486
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98. doi:10.1037/0033-2909.96.1.72
- 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. doi:10.1037/0021-9010.91.3.594
- Lanning v. SEPTA* (1999), 181 F.3d 478
- McClelland, D.C. (1993). Intelligence is not the best predictor of job performance. *Current Directions in Psychological Science*, 2, 5–6. doi:10.1111/1467-8721.ep10770447
- Murphy, K. R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance*, 15, 173-186. doi:10.1207/s15327043hup1501&02\_11
- Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance*, 15, 161-171. doi:10.1080/08959285.2002.9668089
- Patterson v. McLean Credit Union* (1988), 491 U.S. 164

- Ployhart, R. E. & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*, 153-172. doi:10.1111/j.1744-6570.2008.00109.x
- Price Waterhouse v. Hopkins* (1989), 490 U.S. 228
- Schmidt F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187–210.  
doi:10.1080/08959285.2002.9668091
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274. doi:10.1037/0033-2909.124.2.262
- Shoenfelt, E. L., & Pedigo, L. C. (2005). A review of court decisions on cognitive ability testing, 1992-2004. *Review of Public Personnel Administration*, *25*, 271-287. doi: 10.1177/0734371X05276800
- Sternberg, R. J., & Wagner, R. K. (1993). The g-ocentric view of intelligence and performance is wrong. *Current Directions in Psychological Science*, *2*, 1–5.  
doi:10.1111/1467-8721.ep10770441
- Soete, B. D., Lievens, F., & Druart, C. (2013). Strategies for dealing with the diversity-validity dilemma in personnel selection: Where are we and where should we go? *Journal of Work and Organizational Psychology*, *29*, 3-12. doi:10.5093/tr2013a2
- Spearman, C. (1904). “General intelligence,” objectively determined and measured. *The American Journal of Psychology*, *15*, 201-292. doi:10.1037/11491-006

- Terpstra, D. A., Mohamed, A. A., & Kethley, R. B. (1999). An analysis of federal court cases involving nine selection devices. *International Journal of Selection & Assessment*, 7, 26. doi:10.1111/1468-2389.00101
- Wards Cove Packing Co. v. Antonio* (1989), 490 U.S. 642
- Werner, J. M., & Bolino, M. C. (1997). Explaining US courts of appeals decisions involving performance appraisal: Accuracy, fairness, and validation. *Personnel Psychology*, 50, 1-24. doi:10.1111/j.1744-6570.1997.tb00898.x
- Williams, K. Z., Schaffer, M. M., & Ellis, L. E. (2013). Legal risk in selection: An analysis of processes and tools. *Journal of Business and Psychology*, 28, 401-410. doi:10.1007/s10869-013-9299-4

## Appendix A: Coding Factors

### *Coding Factors*

Coding Factor	Definition
Court Level	Was the case held in the district or circuit court?
Gender of plaintiff	Is the plaintiff male or female? (NA indicates gender was an irrelevant factor)
Basis for lawsuit (claim)	What does the plaintiff argue as the basis for discrimination? (race, gender, age, etc.)
Industrial, professional, or civil service work	Is the job in question in the lawsuit one that is industrial, professional, or civil service work?
Class action or individual plaintiff	Is the plaintiff one person or a class action?
Standardized and/or professionally developed cognitive ability test	Was the cognitive ability test that was used in the selection process professionally developed or not?
In-house or consultant	Was the cognitive ability test in question developed in-house or by a consultant?
Was the test validated?	Was the cognitive ability test that was used in the selection process validated?
Were other tests involved?	Did the selection process include other tests, such as a physical endurance test?
Jury	Was a jury present during the court proceedings?
Finding	Did the court rule in favor of the defendant or the plaintiff?
Cut off scores	Was the cutoff score for selection set arbitrarily or by some objective standard?

Note: The factor coding for each of the cases was completed as follows: Court Level: 1=District, 2=Circuit; Claim: 1=Race, 2=Gender; Class Action or Individual: 1=Individual, 2=Class Action; Work: 1=Industrial, 2=Civil Service, 3=professional; Professionally Developed: 1=No, 2=Yes; Developed by: 1=In-House, 2=Consultant; Validated: 1=No, 2=Yes; Other Tests: 1=No, 2=Yes; Jury: 0=Not Present, 1=Present; Finding: 1=Plaintiff, 2=Defendant; Cutoff Score: 1=No, 2=Yes.

## Appendix B: List of Court Cases Included

- Adams v. City of Chicago*, 135 F.3d 1150 (1997); 1998 U.S. App. LEXIS 1477
- Allen v. City of Chicago*, 351 F.3d 306 (2003); 2003 U.S. App. LEXIS 24677
- AMAE v. State of California*, 231 F.3d 572 (2000); 2000 U.S. App. LEXIS 27040
- Banos v. City of Chicago*, 98 C 7629 (2004); 2004 U.S. Dist. LEXIS 5176
- Bazile v. City of Houston*, 895 F. Supp. 2d 718 (2012); 2012 U. S. Dist. LEXIS 14712
- Bew v. City of Chicago*, 96 C 1488 (2000); 2000 U.S. Dist. LEXIS 5526
- Brown v. City of Chicago*, 917 F. Supp. 577 (1996); 1996 U.S. Dist. LEXIS 5762
- Bryant v. City of Chicago*, 200 F.3d 1092 (1999); 2000 U.S. App. LEXIS 528
- Carrabus v. County of Suffolk*, 119 F.2d 221 (2000); 2000 U.S. Dist. LEXIS  
15845
- Fickling v. New York State Department*, 909 F. Supp. 185 (1995); 1995 U.S. Dist. LEXIS  
19235
- Green v. Town of Hamden*, 73 F. Supp. 2d 192 (1999); 1999 U.S. Dist. LEXIS 19463
- Lewis v. City of Chicago*, 2005 U.S. Dist. LEXIS 42544
- Local Union v. Mississippi Power and Light*, CV 3:96-CV-572WS (2004). 2004 U. S.  
Dist. LEXIS 31182
- MOCHA v. City of Buffalo*, 689 F.3d 263 (2012); 2012 U. S. App. LEXIS 15715
- NAACP v. City of Springfield*, 139 F. Supp. 2d 990 (2001); 2001 U.S. Dist. LEXIS 4753
- Nash v. City of Jacksonville*, 895 F. Supp. 1536 (1995); 1995 U.S. Dist. LEXIS 16397
- Reynolds v. AL Department of Transportation*, 295 F. Supp 2d 1298; 2003 U.S. Dist.  
LEXIS 23987
- Rudder v. District of Columbia*, 95-7169, 95-7170, 95-7171 (1995); 1996 U. S. App.  
LEXIS 30527

*Sanchez v. City of Santa Ana*, CV 79-1818 KN (1995); 1995 U.S. Dist. LEXIS 21224

*Smith v. City of Boston*, CV 12-10291-WGY (2015); 2015 U. S. Dist. LEXIS 154468

*United States of America v. The State of Delaware*, Civ. 01-020-KAJ (2004); 2004

U.S. Dist. LEXIS 4560

*United States of America v. City of New York*, 731 F. Supp. 2d 291 (2010); 2010 U. S.

Dist. LEXIS 78641

*Williams v. Ford Motor Company*, 97-2049, 98-1256 (1998); 1999 U.S. App. LEXIS

18370

*Zottola v. City of Oakland*, 01-15283 (2002); 2002 U.S. App. LEXIS 3596

### **Appendix C: List of Court Cases Excluded**

*Antonelli v. State of New Jersey*, Civ. 00-5725 WHW (2004); 2004 U.S. Dist.

LEXIS 5587

*Brunet v. City of Columbus*, 58 F.3d 251 (1995); 1995 U.S. App. LEXIS 15896

*Gonzales v. Galvin*, 151 F.3d 526 (1997); 1998 U.S. App. LEXIS 17560

*Jordan v. City of New London*, 15 BNA IER CAS 919 (1999); 1999 U.S. Dist. LEXIS  
14289

## Appendix D: Court Case Individual Coding Results

<i>Case</i>	<i>Court Level</i>	<i>Claim</i>	<i>Class Action or Individual</i>	<i>Work</i>	<i>Professionally Developed</i>	<i>Developed by</i>	<i>Validated</i>	<i>Other Tests</i>	<i>Jury</i>	<i>Finding*</i>	<i>Cutoff Score**</i>
<i>Adams v. City of Chicago</i>	Circuit	Race	Class Action	Civil	Yes	In-House	Yes	No	No	Defendant	No
<i>Allen v. City of Chicago</i>	Circuit	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Defendant	No
<i>AMAE v. State of California</i>	Circuit	Race	Class Action	Civil	Yes	In-House	Yes	No	No	Defendant	No
<i>Banos v. City of Chicago</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Defendant	No
<i>Bazile v. City of Houston</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Plaintiff	No
<i>Bew v. City of Chicago</i>	Circuit	Race	Class Action	Civil	Yes	Consultant	Yes	No	No	Defendant	No
<i>Brown v. City of Chicago</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Defendant	No
<i>Bryant v. City of Chicago</i>	Circuit	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Defendant	No
<i>Carrabus v. County of Suffolk</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Defendant	No
<i>Fickling v. New York State Dept.</i>	District	Race	Class Action	Civil	Yes	In-House	No	No	No	Plaintiff	No
<i>Green v. Town of Hamden</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Plaintiff	Yes
<i>Lewis v. City of Chicago</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Plaintiff	Yes
<i>Local Union v. Mississippi Power and Light</i>	District	Race	Class Action	Industrial	Yes	Consultant	Yes	No	No	Plaintiff	Yes
<i>MOCHA v. City of Buffalo</i>	Circuit	Race	Class Action	Civil	Yes	In-House	Yes	No	No	Defendant	No
<i>NAACP v. City of Springfield</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	No	No	Defendant	No
<i>Nash v. City of Jacksonville</i>	District	Race	Individual	Civil	Yes	Consultant	Yes	Yes	No	Defendant	No

\*The coding for finding was reduced to defendant or plaintiff, some were summary judgment.

\*\*The coding for this category may be misleading, for further information check the discussion section regarding each case.

<i>Case</i>	<i>Court Level</i>	<i>Claim</i>	<i>Class Action or Individual</i>	<i>Work</i>	<i>Professionally Developed</i>	<i>Developed by</i>	<i>Validated</i>	<i>Other Tests</i>	<i>Jury</i>	<i>Finding*</i>	<i>Cutoff Score**</i>
<i>Reynolds v. AL Dept. of Transportation</i>	District	Race	Class Action	Civil	Yes	In-House	Yes	Yes	No	Defendant	No
<i>Rudder v. District of Columbia</i>	Circuit	Race	Class Action	Civil	Yes	Consultant	Yes	No	No	Defendant	No
<i>Sanchez v. City of Santa Ana</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Defendant	No
<i>Smith v. City of Boston</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	No	No	Plaintiff	Yes
<i>USA v. City of New York</i>	District	Race	Class Action	Civil	Yes	In-House	No	Yes	No	Plaintiff	Yes
<i>USA v. The State of Delaware</i>	District	Race	Class Action	Civil	Yes	Consultant	Yes	Yes	No	Plaintiff	Yes
<i>Williams v. Ford</i>	Circuit	Race	Class Action	Industrial	Yes	Consultant	Yes	Yes	No	Defendant	No
<i>Zottola v. City of Oakland</i>	Circuit	Race	Individual	Civil	Yes	In-House	Yes	Yes	Yes	Defendant	No

\*The coding for finding was reduced to defendant or plaintiff, some were summary judgment.

\*\*The coding for this category may be misleading, for further information check the discussion section regarding each case.