

East Tennessee State University

Digital Commons @ East Tennessee State University

Undergraduate Honors Theses

Student Works

12-2020

Comparative Views on Age Discrimination Within Appellate Court Decisions: Utilizing Werner and Bolino's Framework

Daniel Trinkle

Follow this and additional works at: <https://dc.etsu.edu/honors>



Part of the [Business Law, Public Responsibility, and Ethics Commons](#), [Labor and Employment Law Commons](#), and the [Performance Management Commons](#)

Recommended Citation

Trinkle, Daniel, "Comparative Views on Age Discrimination Within Appellate Court Decisions: Utilizing Werner and Bolino's Framework" (2020). *Undergraduate Honors Theses*. Paper 609. <https://dc.etsu.edu/honors/609>

This Honors Thesis - Withheld is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.

**Comparative Views on Age Discrimination Within Appellate Court Decisions:
Utilizing Werner and Bolino's Framework**

By

Daniel Jude Trinkle

AAS Northeast State Community College, 2018

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Honors Bachelor of Business Administration
in the
Department of Management and Marketing
ETSU College of Business and Technology

© Daniel Jude Trinkle

East Tennessee State University

Winter 2020

Acknowledgement

First and foremost, I would like to thank my esteemed advisor Dr. Michael McKinney for all of the hard work, time, and patience put into my study. With his invaluable knowledge and experience in the field of study, I have been able to learn and grow academically through this research. Additionally, I would like to thank Dr. Allen Gorman and Dr. Clifford Koen for their technical assistance within my study. I would also like to thank the East Tennessee State Honors College for giving me the opportunity to conduct this research. Lastly, I would like to thank my wife Mara, in addition to all of my other family and friends. Without any of their support I would have never been able to accomplish the research I have conducted.

Comparative Views on Age Discrimination Within Appellate Court Decisions: Utilizing Werner and Bolino's Framework

Introduction

The practice of conducting performance appraisals has always been a major topic of debate for business research (Thomas, Bretz, 1994). In addition to the implications involving company performance, there has been a great deal of discussion regarding the legality of performance appraisals (Feild & Holley, 1982; Freeland, 1993). Some of the most important variables that affect performance appraisal legality include the performance rating, the basis by which the appraisal was conducted, and whether a job analysis was conducted for the position. Additionally, the adverse employment decision involved varies from case to case. Some of these decisions include termination, denied promotion, or denied pay. The legal precedence of performance appraisal has been subject to some discussion. Werner and Bolino (1997) cite the Uniform Guidelines on Employee Selection Procedures (1978), when looking at regulations regarding performance appraisal. According to the UGESP, a performance appraisal is looked upon as an employment "test" within the eyes of the court (1978).

In 1997, a study was conducted by Jon Werner and Mark Bolino titled, "Explaining U.S. Courts of Appeals Decisions Involving Performance Appraisal: Accuracy, Fairness, and Validation (1997)." This study looked upon appellate court cases which involved a performance appraisal. Upon examining these cases, variables were coded to find correlations between certain variables and case outcomes; however, since Werner and Bolino (1997), very few studies of similar nature have been conducted and no studies of the exact same nature have been performed. This study is based on the work of Werner and Bolino (1997), in order to display updated results of the same nature. New data exclusive to this study was also added and evaluated in addition to the previous study's findings.

Werner (2001) defines performance appraisals as, "...a specified period of time, all or a group of employees' work performance, behaviors, or traits are individually rated, judged, or described by a person other than the rated employee and the results are kept by the organization (2001)." According to a more recent article from Wellington Zondi (2017), it is shown that performance appraisal is a multi-dimensional definition, involving factors such as employers placing value on employees, employees receiving proper feedback, and the appraisal acting as an opportunity to establish relationships between both parties. Zondi (2017) utilizes several research studies to come to this conclusion, particularly the works of Coens & Jenkins (2000), Forsyth (2002), Langdon and Osborne (2001) and Hunt (2005).

Some of the biggest variables looked upon within Werner and Bolino's (1997) study involve whether or not the performance appraisal was trait or behavioral oriented. Wiersome, Van Der Berg, and Latham (1997) describe trait based appraisals as measures of, "personality traits, such as commitment, initiative, or dependability." Wiersome, Van Der Berg, and Latham (1997) divide the description of behavioral oriented appraisals into two subcategories: (1) behavioral observation scales (BOS) and (2) behavioral expectation scales (BES). They describe BOS's as, "algebraically summated, Likert-type scales based on critical incident job analysis," while in BES's, "critical incidents are ordered on a continuum to define outstanding, average, and unacceptable performance (1997)." For the sake of this study, both variations of behavioral appraisals are looked at within the same context.

Among the biggest variables looked upon in this study, as well as Werner and Bolino's (1997) study, is the employer's utilization of a job analysis. Safdar, Waheed, and Rafiq (2010) describe a job analysis as, "a human resource management practice is a systematic process for collecting and analyzing information about jobs. Job analysis data is perhaps the most widely

gathered type of organizational information for developing human resource management systems.” Safdar, Waheed, and Rafiq (2010) describe from their studies, as well as other previous studies, that there are clear linkages between overall performance and a proper job analysis being conducted. Werner and Bolino (1997) additionally showed from their study that performance appraisals involving a job analysis had much better outcomes in favor of the employer.

The following study uses the concepts developed by Werner and Bolino (1997) by utilizing the same practices applied within their study. This includes coding many of the same variables, with the addition of a few new variables. The following study also utilizes the same data analysis techniques that were present in Werner and Bolino’s (1997) study. Where this study differs lies within the type of data that is evaluated. The most important facet of data evaluated comes from the form of discrimination involved in each case; meaning, whether the cases were based on race, gender, age, or any other form of discrimination. These specific data points assisted in showing how the courts comparatively looked at each form of discrimination, in spite of the performance appraisal or its rating.

Literature Review

Studies regarding performance appraisal within court cases have been prominent for quite some time. Of course, the main basis for this study derives from nearly all the same standards used within Werner and Bolino’s (1997) study. A major influence on the study of Werner and Bolino (1997) was the work of Feild and Holley (1982), who conducted a similar study prior to their own work. Feild and Holley (1982) showed there were correlations between certain variables and the outcome of the cases. The variables noted for having the strongest correlations include (1) type of organization, (2) provision of written instructions, (3) trait vs. behavioral

oriented appraisal, (4) use of job analysis, (5) and review of appraisal results with employee (Feild and Holley, 1982). While displaying strong correlations between particular variables, the study was limited by its small sample size, only using 66 cases (Feild and Holley, 1982). This is similar to the research of Feild and Thompson, who utilized only 31 cases, as noted by Werner and Bolino (1997). These studies are remarkably similar to Werner and Bolino (1997), as they used data analysis to display correlations between case outcomes and variables related to performance appraisals. Werner and Bolino (1997) did, however, improve upon these studies by looking at a greater number of court cases.

Werner and Bolino (1997) also noted several other research studies that influenced their own works, including the works of Ash & McRae (1985), Martin & Bartol (1991), and Schuster & Miller (1984). Martin & Bartol (1991) were particularly influential, as their study looked into relationships between performance and various outcomes such as promotion, layoff, merit pay, and demotion. Martin and Bartol (1991) looked into greater detail of these variables by dividing them into sub-variables, which showed overall more detail for these specific cases than Werner and Bolino (1997). The study specifically looked at outcomes in different cases with particular criteria such as “Discrimination in Promotion,” “Person Selected Had Superior Credentials,” “Employee Exhibited Serious Shortcomings in Present Job (2000).” While information was limited by fewer cases, more data specific to these particular cases could be gathered as a result.

Since Werner and Bolino (1997), very few similar studies have been conducted. In his assessment of the benefits and costs of performance appraisals, Nickols (2007) cites legal protection as one of the biggest reasons for proper performance evaluations. Hennessey and Bernardin (2003) looked into the potential for adverse impact within performance appraisals, particularly by viewing the differences between criterion-related and non-criterion related

appraisal. Schraeder, Self, & Lindsay (2006) evaluated the differences in rank-order and banding approaches to performance appraisal. While not necessarily looking into quantitative data, Schraeder, Self, and Lindsay (2006) were able to look at this issue from a more qualitative perspective.

Lee, Havighurst, and Rassel (2004), whose work was highly influenced by Werner and Bolino (1997), evaluated differences between circuits. In particular, the research yielded positive results regarding conservative and liberal circuits, highlighting whether conservative courts based decisions on an appraisal's validity and whether liberal courts based decisions on an appraisal's fairness (Lee, Havighurst, Rassel, 2004). Giumetti and Schroeder (2015) evaluated Forced Distribution Rating Systems (FDRS). Their research discovered many links between these performance evaluation techniques and adverse impact (2015).

Hypotheses

The present study utilized a mix of hypotheses from Werner and Bolino (1997), in addition to hypotheses tested exclusively for this study. Werner and Bolino (1997) tested a total of seven hypotheses within their study, all to see how certain aspects in performance appraisals affected the outcome. In turn, Werner and Bolino's (1997) hypotheses were based upon the work of Field and Holley (1982). The hypotheses utilized by Werner and Bolino (1997) attempted to examine whether job appraisals should: (1) be based on a job analysis, (2) emphasize behaviors or results, rather than traits, (3) include specific written instructions to raters, (4) allow review of the appraisal results by employees, (5) involve triangulation, (6) have trained evaluators, and (7) be done frequently.

Out of Werner and Bolino's (1997) seven hypotheses, four were tested for this study. Specific and written instructions to the evaluator were not found within cases; thus, the variables were not accounted for in this study. The same is true for (4) reviewing the appraisal with employee and (6) having trained evaluators. Triangulation was not directly accounted for, as the action of triangulation was split into two variables: whether the appraisal had an accountability mechanism and/or multiple raters.

Hypothesis 1: Decisions in favor of organizations will be more likely when the appraisal system is based on a job analysis.

Hypothesis 2: Decisions in favor of organizations will be more likely when the appraisal is behavior- or results- oriented.

Hypothesis 3: Decisions in favor of employer will be more likely when there is triangulation among multiple raters.

Hypothesis 4: Decisions in favor of organizations will be more likely when formal evaluations are conducted more frequently.

The next set of hypotheses were developed exclusively for this study in an attempt to find differences between how the different forms of discrimination are viewed in court. The study looks upon an employee's rating in order to view if the appraisal rating had any effect on the outcome. The overall assumption to all of the hypotheses is that courts mostly favor organizations within age discrimination cases, whether or not the employee received positive reviews. This includes comparison with other specific forms of discrimination such as race or gender; however, it also includes looking upon age discrimination in comparison to all other forms of discrimination combined together. In summary, the hypotheses state that cases involving age discrimination are easier for the organization to win when an employee has a

positive or average performance appraisal in comparison to (1) racial discrimination, (2) gender discrimination, and (3) all non-age discrimination cases. Following this approach, it will also be hypothesized that it is easier for the employee to win with a negative performance appraisal in cases involving (1) racial, (2) gender, or (3) any non-age discrimination. If this holds true, then racial and gender discrimination should be indifferent from each other.

Hypothesis 5: Cases involving age discrimination will favor the organization, in spite of appraisal rating, in comparison to racial discrimination cases.

Hypothesis 6: Cases involving age discrimination will favor the organization despite appraisal rating in comparison to gender discrimination cases.

Hypothesis 7: Cases involving age discrimination will favor the organization despite appraisal rating in comparison to all non-age discrimination cases.

Hypothesis 8: Cases involving age discrimination will favor the organization when employee has a positive or average performance rating in comparison to all non-discrimination cases.

Hypothesis 9: Cases involving non-age discrimination will favor the employee when employee has a negative performance rating in comparison to age discrimination cases.

Hypothesis 10: Cases involving racial discrimination will favor the employee equally in comparison to gender discrimination cases in spite of performance appraisal rating.

All of the listed hypotheses attempt to test the same basic concept, which indicates that age discrimination is much more difficult to win in court, despite the presence of performance appraisal ratings. Hypotheses 5-7 look at the overall scope of age discrimination cases, whether

the employee had a positive, average, or negative review. Hypotheses 8-9 serve the same purpose, but further examine whether positive appraisals have an effect on age discrimination cases or whether negative performance appraisals have an effect on racial, gender, or all non-age discrimination cases. Hypothesis 10 affirms the statements from the first nine hypotheses; after all, where differences lie between age discrimination and all other forms of discrimination, there should also be similarities between the opposing forms of discrimination.

Methodology

To maintain continuity with the Werner and Bolino (1997) study, the methodologies utilized in this study aligned very closely to their own. This includes areas such as selecting legal cases and coding variables; however, there is some differentiation, particularly in the some of the variables accounted for in the study. For this particular study, different data points were evaluated in comparison to what was evaluated within Werner and Bolino's (1997) study; thus, additional variables were added.

Selecting Court Cases

The methodology for selecting cases was conducted nearly identical to Werner and Bolino (1997). Similar to Werner and Bolino (1997), Westlaw was utilized for the study when selecting cases. Cases were selected from the years 1996-2020. Only cases from the United States Courts of Appeals were chosen for this study. While there were certainly cases from both the district courts and United States Supreme Court, solely utilizing circuit courts of appeals cases was done for the following reasons: (1) there is greater legal precedence within the courts of appeals than the district courts, (2) it is much easier to test differences in patterns across 13 courts of appeals as opposed to the numerous district courts, (3) utilizing the courts of appeals

narrows down the searching criterion and allows for a higher percentage of its cases to be used (Werner and Bolino, 1997).

Search queries typically involved the terms “performance appraisal” and its subsidiaries such as “performance review” or “performance evaluation.” This utilization allowed for a simpler method of finding cases where performance reviews were brought to the court. Typically paired with these terms was the word “discrimination,” as this eliminates cases that were commercial or contract law. One hundred court cases, taking place from 1996-2020, were selected for this particular study. While many other cases were viewed, the cases that were selected were chosen because they (1) involved a performance appraisal(s), and (2) had enough information regarding the appraisal to code the variables considered for the study.

Variables

As with selecting cases, the coding of variables was performed almost exactly as Werner and Bolino (1997); however, there is slight differentiation within the variables. This is due to some variables in Werner and Bolino’s (1997) study not necessarily providing enough detail. A major example of this practice includes Werner and Bolino (1997) categorizing job analysis under two categories: 0 = no job analysis, 1 = job analysis. This practice does not consider that certain cases may have just excluded the fact that there was a job analysis; after all, just because a case does not mention it, does not necessarily mean a job analysis was never conducted. Within this study a third variable was added, which was labeled as such: 2 = Unannounced.

Werner and Bolino (1997) describe the dependent variable of the study as the decision reached by the court. Here, the variables were coded as 0 if a case went in favor of the plaintiff and coded as 1 if a case went in favor of the defendant. While most cases were clearly either for

the plaintiff or defendant, some cases were a little harder to decide. Many cases were modified and granted both parties certain allowances. Additionally, certain cases were remanded for retrial, meaning the future was uncertain for the final decision. For this study, in order to simplify the results, the decision to code for the plaintiff or defendant relied solely on the appellate court decision; meaning, if a case was modified or remanded, it would be coded in favor of the losing party from the trial court's decision. This is due to the fact that if a party loses a case in trial, the success relies solely on the appellate court granting any sort of relief to the losing party.

Variables containing general information coded for the study included ones that described the company: private = 1; public = 0. Unlike Werner and Bolino (1997), this study also accounted for government, university, and non-profit organizations: government = 3; university = 4; non-profit = 5. Companies were also coded for size on a scale of 1 to 4: 1 = Small (< 100 employees), 2 = Medium (100-500 employees), 3 = Medium-Large (500-1000 employees), and 4 = large (> 1000 employees). Regarding the characteristics of the plaintiff, major factors coded were ethnicity and gender. Location was also considered, particularly the court system in which the trial took place. The appellate court location was identified, in addition to the district court from which the case originated. In terms of general information, the last variable coded was the year the appellate court case took place.

Information that describes the performance appraisal in more detail were considered. The appraisal's rating was coded as such: positive = 1; average = 2; negative = 3. Time between performance appraisal was coded identically to Werner and Bolino (1997) ranging from: 0 = less often than once a year, 1 = once a year, 2 = six months to a year, 3 = three months to six months, and 4 = more often than every three months. The basis of the performance review was coded depending on if the appraisal was trait, behavioral, or a less traditional method such as

management by objective. More specific information regarding the performance appraisal included information regarding whether there were multiple evaluators, measures of counterproductive behaviors, or an accountability mechanism.

Variables accounting for the type of discrimination were among the most important factors evaluated in the study. These factors were coded as such: Race = 1, Gender = 2, Age = 3, Religious = 4, Disability = 5, National Origin = 6, Multiple = 7. An additional variable was added for this particular study, which showcased exclusively whether or not the case was based on age discrimination: Age Discrimination = 1, Non-Age Discrimination = 0. For the sake of this study, age discrimination was substantially evaluated; thus, the added variable was necessary to compare age discrimination to all the other variables combined. While age discrimination was looked upon in comparison to specific types of discrimination, this variable allowed for the analysis of how age discrimination generally compared to the other forms of discrimination combined.

The last set of variables is related to information about the case itself. Whether the case was based on disparate impact or disparate treatment was evaluated. The form of adverse employment decision was a factor evaluated for the study. Relevant information regarding the specific outcome of the case was coded: affirmed = 1; reversed = 2; modified = 3; remanded = 4. Whether or not the case was a class action was additionally evaluated and coded.

Procedure

Information was collected by viewing characteristics written within each case. The information from each case was then coded for every relevant variable. Certain cases required additional information; thus, the district court cases were evaluated in such cases. Information

regarding the company required looking to sources outside of Westlaw. This information was typically on company websites; however, missing information was mostly found through Bloomberg. Bloomberg was used especially in the determination of whether a company operated publicly or privately.

Data Analysis

To maintain strong continuity with Werner and Bolino (1997), similar data analysis was conducted. As with Werner and Bolino (1997), univariate chi-square analysis was applied to test the hypotheses and understand correlation between variables. This allowed the study to see whether there were strong correlations between variables and, if so, see if the correlation is significant enough to prove anything. Chi-square tests were done on four of Werner and Bolino's (1997) hypotheses, in addition to the added hypotheses for the sake of this study. A difference in chi-square analysis was conducted when comparing the different data sets for hypotheses 5-10.

Data

Judicial decisions highly favored defendants within the cases. In total, 79 cases (79%) went in favor of the defendant, 75 of which were simply affirmed decisions from the trial court level. Cases came from all 12 circuits in the courts of appeals, with the most prominent being the 7th Circuit (16) and the least prominent being the District of Columbia Circuit (3). When looking at the classifications of discrimination, the majority of the cases dealt with age discrimination (43%). The rest of the cases were distributed as follows: race discrimination (26%), gender discrimination (18%), disability discrimination (3%), national origin discrimination (2%). Cases involving multiple counts of discrimination accounted for 8% of the cases.

Most of the cases were not part of class action lawsuits, as these types of cases only made up 8% of the cases used. Of the selected cases, 54 involved male plaintiffs, 44 involved female plaintiffs, and 2 cases involved parties of multiple genders. Ethnicity was more difficult to categorize, as in most instances only the cases involving racial discrimination explicitly stated the ethnicity of the plaintiff; thus, 63% of the cases showed no evidence one way or another regarding the specific ethnicity of the plaintiff. For the cases where ethnicity was stated, the majority of cases involved African American plaintiffs (70.27%).

Behavioral based appraisals outweighed trait based appraisals by a substantial amount, as 62 cases had behavioral appraisals while only 7 cases had trait appraisals. Three of the cases did have a management-by-objectives (MBO) based appraisal, and 9 of the cases lacked enough information to determine any conclusion. One interesting data point to note was the variable examining evidence of a job analysis. Forty-five of the cases showed evidence that the appraisals had evidence of a job analysis, while 55 of the cases did not. For those appraisals where a job analysis was conducted, all but one of the cases went in favor of the employer (97.78%). Even more interesting in cases where there was no evidence of a job analysis, only 63.64% of the cases were won by the employer.

Hypothesis Testing

Out of Werner and Bolino's (1997) four tested hypotheses only the job analysis hypothesis (Hypothesis 1) proved accurate ($\text{Chi Stat} = 17.38 > \text{Chi Critical} = 3.84$). This further shows that job analyses are a critical factor in court outcomes. From the hypotheses testing, neither behavioral based appraisals (Hypothesis 2), appraisals with triangulation (Hypothesis 3), nor more frequently performed appraisals (Hypothesis 4) had any statistical difference. Neither sub-variable for triangulation, appraisals with multiple raters nor accountability mechanisms

tested statistically significant. Appraisal frequency approached being statistically significant; however, it still came up just short of the critical value ($\text{Chi Stat} = 3.12 < \text{Chi Critical} = 3.84$). The lack of statistical significance is more than likely due to fewer data points in comparison to the Werner and Bolino (1997) study.

Age Discrimination Comparison

From a descriptive statistical standpoint, there are clear differences between the type of discrimination and outcome when factoring in the performance rating. Out of the 43 age discrimination cases, 6 (13.95%) were won by the plaintiff and 37 (86.05%) were won by the defendant. Respectively, out of all the racial discrimination (26) and gender discrimination (18) cases, 18 (69.23%) and 13 (72.22%) were won by the defendant; however, it is difficult to draw a significant conclusion, considering the difference in the number of cases for each type of discrimination. Evaluating a relatively equal number of cases compared to age discrimination, the data already shows some discrepancies between age discrimination cases and non-age discrimination cases (age discrimination = 86.05% > non-age discrimination = 73.68%)

Adding in data regarding performance rating shows further significance, as there are a few major points to note. Of the 37 age discrimination cases won by the plaintiff, 18 (48.65%) had positive performance reviews, 9 (24.32%) had average reviews, and 10 (27.03%) had negative reviews. It is interesting to note that in the majority of cases where the defendant won, the plaintiff had an overall positive review; furthermore, these cases often went in favor of the defendant when the employer had either an average review or positive review (72.97%). It is also interesting to note that out of the age discrimination cases where the plaintiff won, the plaintiff never had a negative review.

Once again, this data is relatively meaningless without comparison to other specific forms of discrimination. In terms of racial discrimination cases (61.11%) and gender discrimination cases (53.85%), a much higher percentage of defendants won when the plaintiff had a negative review; meaning, there is a closer linkage to the rating and outcome than with age discrimination. Viewing all non-age discrimination cases where the defendant won, 26 (60.47%) had negative reviews. To further note, in all non-age discrimination cases where the plaintiff won, none recorded a positive review and 10 (66.67%) had negative reviews.

To further establish these differences, hypotheses testing was required. It is important to note that all of the hypotheses tested exclusively for this study (Hypotheses 5-10) yielded positive results. Before the added hypotheses were tested, differences were tested between the different forms of discrimination, without accounting for the performance appraisal rating. As with Werner and Bolino (1997), this test proved insignificant ($\text{Chi Stat} = 3.53 < \text{Chi Critical} = 7.82$). Utilizing a difference in chi-square analysis, there were enough variations to accept the alternative hypotheses in hypotheses 5-9. Unique from hypotheses 5-9, hypothesis 10 showed no statistical significance between racial discrimination and gender discrimination. With the lack of statistical difference displayed in the chi-square analysis ($\text{Chi Stat} = 0.65 < \text{Chi Critical} = 5.99$), the null hypothesis was unable to be rejected.

Hypothesis 5	<p>H0 = There is no difference between performance rating and outcome when comparing age and racial discrimination</p> <p>H1 = There is a difference between performance rating and outcome when comparing age and racial discrimination</p> <p><u>Statistical Significance:</u> Chi Stat (7.62) > Chi Critical (5.99) Reject H0</p>
--------------	---

Hypothesis 6	<p>H0 = There is no difference between performance rating and outcome when comparing age and gender discrimination</p> <p>H1 = There is a difference between performance rating and outcome when comparing age and gender discrimination</p> <p><u>Statistical Significance:</u> Chi Stat (8.74) > Chi Critical (5.99) Reject H0</p>
Hypothesis 7	<p>H0 = There is no difference between performance rating and outcome when comparing age discrimination and non-age discrimination</p> <p>H1 = There is a difference between performance rating and outcome when comparing age discrimination and non-age discrimination</p> <p><u>Statistical Significance:</u> Chi Stat (6.67) > Chi Critical (5.99) Reject H0</p>
Hypothesis 8	<p>H0 = There is no difference regarding negative reviews in outcome between age discrimination and non-age discrimination</p> <p>H1 = There is a difference regarding negative reviews in outcome between age discrimination and non-age discrimination</p> <p><u>Statistical Significance:</u> Chi Stat (9.25) > Chi Critical (3.84) Reject H0</p>
Hypothesis 9	<p>H0 = There is no difference regarding positive reviews in outcome between age discrimination and non-age discrimination</p> <p>H1 = There is a difference regarding positive reviews in outcome between age discrimination and non-age discrimination</p> <p><u>Statistical Significance:</u> Chi Stat (15.56) > Chi Critical (7.82) Reject H0</p>
Hypothesis 10	<p>H0 = There is no difference between performance rating and outcome when comparing racial and gender discrimination</p> <p>H1 = There is a difference between performance rating and outcome when comparing racial and gender discrimination</p> <p><u>Statistical Significance:</u> Chi Stat (0.65) < Chi Critical (5.99) Accept H0</p>

Discussion

Despite fewer cases there were still many similarities to Werner and Bolino's (1997) work. This is much in the same way Werner and Bolino's (1997) study closely linked with Field and Holley (1982). Nonetheless, there were some major differences between the two studies. When comparing descriptive statistics, court outcomes were particularly different within this study. Little over half of Werner and Bolino's (1997) outcomes went in favor of the defendant (58.6%), whereas a majority of cases went in favor of the defendant in this study (79%). Additionally, it is interesting to view the amount of trait appraisals within this study in comparison to Werner and Bolino (1997). Considering their study had 45% of appraisals listed as trait based, it is interesting to note the lack of trait based appraisals in this study (7%).

Many of the other findings from this study resemble the findings from Werner and Bolino's (1997) study. As with Werner and Bolino (1997), the hypothesis regarding job analysis (Hypothesis 1) was found to be supported. Additionally, the appraisal basis (Hypothesis 2) and appraisal frequency (Hypothesis 3) did not have a statistical relation; thus, failing to support this hypothesis in the same way as Werner and Bolino (1997). As previously mentioned, however, appraisal frequency was very close to having statistical significance, indicating that further tests must be performed with more cases to see if there is significance in this area. This is similar to appraisal basis, which had far too few trait-based appraisals to really compare significance with behavioral based appraisals. The only hypothesis not found similar to Werner and Bolino (1997) involved triangulation (Hypothesis 4). In their study, triangulation proved to be statistically related to court outcome; unfortunately, looking at both measures by which triangulation was evaluated upon in this study (multi-rater and accountability mechanism), neither had a statistical relation with court outcome. Failure to support this hypothesis could be due to the following

reasons: (1) dividing this variable into two sub-variables, (2) the lack of cases in comparison to Werner and Bolino (1997).

Still, these differences are mostly subtle and hardly significant. With the exception of behavioral and trait comparison, as well as the failure to prove Hypothesis 4, there is little to dispute the results of Werner and Bolino's (1997) work. The lack of trait-based appraisals is more than likely due to the lack of popularity of these appraisal types when comparing the timeframe of this study to the timeframe utilized by Werner and Bolino (1997). The importance of a job analysis particularly holds true throughout Werner and Bolino's (1997) research, as well as Feild and Holley's (1982) research. It appears that between these three separate works, there is a lack of an overall major change in the data. Indicating that there are only incremental changes throughout these studies across very different timeframes. Even displaying how each study has little variation from one another is a particularly important note of the research conducted.

When discussing the added hypotheses regarding the different forms of discrimination, there was obviously some important data added to Werner and Bolino's (1997) base framework. Further showcasing the similarities in data, the form of discrimination proved statistically insignificant without accounting for the performance rating. Statistical differences were however found between age discrimination in comparison to racial discrimination, gender discrimination, and all non-age discrimination cases, when adding in the variable regarding performance rating. These particular variables were not evaluated or tested by Werner and Bolino (1997).

Further adding to the study, a determination had to be made on whether or not there was a particular difference in plaintiffs with positive/average ratings and plaintiffs with negative ratings. These differences were tested for individually by separating positive/average reviews from negative reviews. Once again, the results showed clear statistical differences which proved

each hypothesis. Paired with the descriptive statistics, there are certain conclusions that come as a result: (1) None of the outcomes for aforementioned forms of discrimination particularly align with their ratings, (2) it is much easier for a defendant to win in a case involving age discrimination when the plaintiff has a positive rating in comparison to non-age discrimination cases, and (3) it is much easier for the plaintiff to win in a case involving non-age discrimination when the plaintiff has a negative review in comparison to age discrimination cases.

Conclusion

The overall goal of this study sought to accomplish two purposes: (1) establish updated data since Werner and Bolino's (1997) study and (2) utilize Werner and Bolino's (1997) framework to discover new evidence regarding age discrimination. Both purposes were met equally well, as hypothesis testing showed similarities between Werner and Bolino's (1997) study, as well as showing statistical significance for newer data sets. The updated data showed there are still certain factors in performance appraisals that are viewed positively by the courts, particularly the presence of a job analysis. The differences that were found were certainly distinct, yet minor when looking at the overall framework. While certainly further testing should be accomplished, the new data gathered from this study shows there are discrepancies in how the court views different forms of discrimination. While these discrepancies may not be distinct, they are still worth noting as the statistical differences are clearly present. Additionally, this new data supports how much more can be done with Werner and Bolino's (1997) framework. The data of Werner and Bolino (1997) requires constant updating, as there is still valuable data yet to be gathered from studies of this sort.

Nonetheless, there are still limitations, both in Werner and Bolino's (1997) study, as well as this one. Werner and Bolino's (1997) own biggest concern regarded the amount of missing

information, which proved true in this study as well. Certain cases do not provide all information needed to test every variable; in fact, most cases rarely have enough information to fill every data cell. Werner and Bolino (1997) also mention that federal district and state court cases also need to be evaluated specifically to have even more understanding of the subject matter. For this particular study, the most prominent issue was the number of cases evaluated. This issue specifically was problematic when comparing data with Werner and Bolino (1997). This problem did not particularly hinder the new data that was gathered; however, more cases could further prove the hypotheses stated in the research. Not mentioned by Werner and Bolino (1997), is the subjectivity of certain variables. For instance, certain cases may have stated that an employee had negative performance rating, but the rating would actually be a 3 or a 3.5 out of 5. By all regards this is at least an average rating; thus, it is up to the person coding the variables to decide on whether to code this as negative, average, or positive.

With this information, it is important to build upon the foundation set by Werner and Bolino (1997), who in turn formed their basis from the work of Feild and Holley (1982). This information is crucial for employers, as they need to understand the facets of an appraisal that hold up the greatest in court. Utilizing this research, employers could avoid potential lawsuits or, at least, strengthen their position when a performance appraisal is at issue in the case. It is important also for researchers to continue to work on this framework, in order to continually update information on the topic. Additionally, researchers can take this framework and find unique uses with its information, such as the study presented here. This study contributes to the already established literature regarding court outcomes and performance. In addition, this study will provide a further basis for studies similar to this one in the future.

References

- Feild, H. S., Holley, W. H., (1982). The relationship of performance appraisal system characteristics to verdicts in selected employment discrimination cases. *Academy of Management Journal (pre-1986)*, 25(2), 392. Retrieved from <https://search.proquest.com/abicomplete/docview/229612326/A55401B71A5E4E02PQ/5?accountid=10771>
- Freeland, M., (1993). Performance appraisal and disciplinary action: the case for. *International Labor Review*, 132(4), 493. Retrieved from <https://search.proquest.com/abicomplete/docview/224006418/8C711B287C8E4E64PQ/40?accountid=10771>
- Giumetti, G. W., Schroeder, A. N., Switzer, F. S., (2015). Forced distribution rating systems: when does “rank and yank” lead to adverse impact? *Journal of Applied Psychology*, 100(1), 180-193. Retrieved from https://psycnet.apa.org/fulltext/2014-24621-001.pdf?auth_token=07e7c4a9d6d936150fc40630b98accdf55e6f76d
- Hennessey, H. W. Jr., Bernardin, J. H., (2003). The relationship between performance appraisal criterion specificity and statistical evidence of discrimination. *Human Resource Management*, 42(4), 143-148. Retrieved from <https://search.proquest.com/abicomplete/docview/222129191/B3418918AB524EA0PQ/13?accountid=10771>
- Lee, J. A., Havighurst, L. C., Rassel, G., (2004). Factors related to court references to performance appraisal fairness and validity. *Public Personnel Management*, 33(1), 61-77. Retrieved from

<https://search.proquest.com/abicomplete/docview/215930429/B3418918AB524EA0PQ/1?accountid=10771>

Martin D. C., Bartol, K. M., (1991). The legal ramifications of performance appraisal: an update. *Employee Relations Law Journal*, 17(2), 257. Retrieved from <https://search.proquest.com/abicomplete/docview/194229268/fulltextPDF/28B9E23A93734DF7PQ/24?accountid=10771>

Nickols F., (2007). Performance appraisal: weighed and found wanting in the balance. *Journal of Quality and Participation*, 30(1), 13-16, 47. Retrieved from <https://search.proquest.com/abicomplete/docview/219109060/CF331BBA8E994CCFPQ/4?accountid=10771>

Safdar, R., Waheed, A., Rafiq, K. H., (2010). Impact of job analysis on job performance: analysis of a hypothesized model. *Journal of Diversity Management*, 5(2), 17-36. Retrieved from <https://search.proquest.com/abicomplete/docview/506643513/96808DFB511C4C8APQ/5?accountid=10771>

Schraeder, M., Self, D. R., Lindsay, D. R., (2006). Performance appraisals as a selection criterion in downsizing: A comparison of rank-order and banding approaches. *Managerial Law*, 48(5), 479-494. Retrieved from <https://search.proquest.com/abicomplete/docview/196365908/B3418918AB524EA0PQ/7?accountid=10771>

Thomas, S. L., Bretz R. D., (1994). Research and practice in performance appraisal: evaluating employee performance in America's largest companies. *S.A.M. Advanced Management*

Journal, 59(2), 28. Retrieved from

<https://search.proquest.com/abicomplete/docview/231132233/D1169A3AC62C4CB7PQ/9?accountid=10771>

Werner J. M., (2001). Abolishing performance appraisals. *Personnel Psychology*, 54(4), 1030-1034. Retrieved from

<https://search.proquest.com/abicomplete/docview/220150533/43F379B05A74527PQ/3?accountid=10771>

Werner, J. M., Bolino, M. C., (1997). Explaining U.S. courts of appeals decisions involving performance appraisal: accuracy, fairness, and validation. *Personnel Psychology*, 50, 1-24.

Wiersome, U. G., Van Der Berg, P. T., Latham, G. P., (1997). Dutch reactions to behavioral observation, behavioral expectation, and trait Scales. *Group Organization Studies*, 20(3), 297. Retrieved from

<https://search.proquest.com/abicomplete/docview/232769623/F0B64BB2B3504E8EPQ/4?accountid=10771>

Zondi, W. B., (2017). The concept of performance appraisal and requirements for its successful implementation. *International Journal of Economic Perspectives*, 11(2), 638-644.

Retrieved from

<https://search.proquest.com/abicomplete/docview/2038221636/B8C1146E7B9B4517PQ/2?accountid=10771>