THE ROLE OF PERFORMANCE APPRAISAL IN LITIGATION:

AN UPDATED ANALYSIS OF CASE LAW

A Thesis

by

ANJELICA M MENDOZA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Stephanie C. Payne
Committee Members,	Ramona L. Paetzold
	Winfred Arthur, Jr.
Head of Department,	Heather C. Lench

December 2020

Major Subject: Psychology

Copyright 2020 Anjelica M Mendoza

ABSTRACT

Performance appraisal (PA) is a cornerstone of human resource management as it is the primary way organizations document and track how well personnel are performing on the job. Despite their widespread utilization, there have been a growing number of calls to eliminate PAs from business practices. This is potentially problematic as PAs are often the only documentation an organization has of an employee's performance. Conceivably, without such documentation, it would be difficult for an organization to legally defend personnel decisions. The purpose of this study is to provide an updated review of the role of PA in employment discrimination litigation. This review examines the extent to which five PA content recommendations (e.g., PA based on a job analysis) and nine PA process-based recommendations (e.g., standardized procedures across employees) for organizations appeared in 462 U.S. Federal Courts of Appeals cases within the last five years (2014-2018). While the recommendations are well established in the research literature, the study found that all of the content recommendations and many of the process recommendations did not appear or were seldom mentioned within the case narratives. Only four of the previously supported process recommendations were associated with an increased likelihood the courts will find in favor of the organization: the use of multiple raters, more performance documentation, the opportunity for the employee to review the ratings, and the opportunity to correct performance deficiencies. An expanded examination beyond Title VII reveals that the previously supported PA recommendations may not be as impactful in current discrimination claims. Analyses of several previously unexamined hypotheses also demonstrate that an organization's sector and the presence of satisfactory PA evidence are related to the likelihood of case decisions. Results have significant practical implications concerning what

organizations should do to maximize the legal defensibility of their decisions. Limitations of the study and recommendations for future examinations are discussed.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by the student's thesis committee chair, Professor Stephanie C. Payne of the Department of Psychological and Brain Sciences with Professor Winfred Arthur, Jr. of the Department of Psychological and Brain Sciences, Professor Ramona L. Paetzold of the Department of Management as committee members. Professor Stan Malos of the School of Management at San Jose State University also provided advice on an earlier version of this project. Coding of case content was completed with assistance from undergraduate research assistants; James Tanner Allen, Addison Dickey, Aileen Dowden, Meagan Elmer, Bailey Lockaby, Wheeler Nakahara, Hannah O'Malley, Dillon Osgood, Cassidy Springer, and Sierra Stryker.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

NOMENCLATURE

Statutes

ADA	Americans with Disabilities Act (1990)
ADEA	Age Discrimination in Employment Act (1967)
CRA	Civil Rights Act of 1964
EPA	Equal Pay Act (1963)
PDA	Pregnancy Discrimination Act (1978)
Section 1981	The Civil Rights Act of 1866, § **1981,** 42 U.S.C.
Title VII	Civil Rights Act of 1964, § **7**, 42 U.S.C. §2000 et seq (1964)

Legal

Courts of Appeals	Federal U.S. Circuit Courts of Appeals
EEOC	Equal Employment Opportunity Commission

Performance Management

HRM	Human Resource Management
I-O	Industrial-Organizational
PA	Performance Appraisal
PIP	Performance Improvement Plan
PM	Performance Management

TABLE OF CONTENTS

Page

ABSTRACT ... ii

CONTRIBUTORS AND FUNDING SOURCES ... iv

NOMENCLATURE ... v

TABLE OF CONTENTS ... vi

LIST OF FIGURES ... viii

LIST OF TABLES ... ix

1. INTRODUCTION ... 1

2. PERFORMANCE APPRAISAL AND LEGAL CONSIDERATIONS ... 3

   2.1. Purpose of Performance Appraisals ... 3
   2.2. Legal Defensibility of Performance Appraisals ... 5
      2.2.1. Previous Examinations of Case Law ... 6
      2.2.2. Performance Appraisal Content ... 15
      2.2.3. Performance Appraisal Process ... 17

3. PERFORMANCE APPRAISAL IN LITIGATION ... 27

   3.1. Legal Discrimination Protections ... 28
   3.2. Litigation Process ... 30
   3.3. Equal Employment Opportunity Commission ... 32
      3.3.1. EEOC Represented Cases ... 34
      3.3.2. Retaliation ... 35
   3.4. Public and Private Sector Organizations ... 36
   3.5. Evidence of Satisfactory Performance ... 37
   3.6. Examples of Performance Appraisals in Discrimination Cases ... 38

4. METHOD ... 42

   4.1. Case Selection ... 42
   4.2. Coding Variables of Interest ... 44
      4.2.1. Discrimination Claims ... 45
      4.2.2. Performance Appraisal Content ... 46
      4.2.3. Performance Appraisal Process ... 47
      4.2.4. Novel Variables ... 48
   4.3. Analysis ... 49

5. RESULTS 51

5.1. Claim Information..51
5.2. Employment-Related Information 52
5.3. Hypothesis Testing..54
5.3.1. Hypothesized Decisions for Employees 62
5.4. Exploratory Analysis 65

6. DISCUSSION...70

6.1. Malos's Recommendations...71
6.2. Comparisons to Previous Examinations 72
6.2.1. Number of Cases Examined..75
6.2.2. Missing Data 76
6.2.3. Possible Explanations of Differences 77
6.3. Contributions...81
6.3.1. The Employee's Perspective..83
6.3.2. Decisions for the Employee. 85
6.3.3. Performance Improvement Opportunities...86
6.4. Limitations 88
6.4.1. Representation of Discrimination Claims. 88
6.4.2. Judicial Bias 89
6.4.3. Case Decisions 90
6.5. Future Directions 90

7. CONCLUSION...93

REFERENCES...95

APPENDIX A 109

APPENDIX B 115

APPENDIX C 117

LIST OF FIGURES

Page

Figure 1 Simplified Equal Employment Opportunity Commission Process 33

Figure 2 Case Identification Flow Chart.. 43

LIST OF TABLES

Page

Table 1 Prior Quantitative Examinations of Performance Appraisal Characteristics in Litigation 8

Table 2 Prior Qualitative Examinations of Performance Appraisal Characteristics in Litigation 10

Table 3 Cases and Recommendations from Previous Examinations of Performance Appraisal in Case Law 12

Table 4 Table of Hypotheses 40

Table 5 Frequency and Decisions of Claims Argued in Appeal 53

Table 6 Logistic Regression Analyses for the Full Sample of Cases 56

Table 7 Logistic Regression Analyses for Title VII Claims 57

Table 8 Summarized Results of Tested Hypotheses 64

Table 9 Exploratory Logistic Regression Analyses 69

Table 10 Summary of the Examined Performance Appraisal (PA) Recommendations 73

Table 11 Comparisons to Previous Quantitative Examinations 74

1. INTRODUCTION

“Exceeds expectations.” “Needs improvement.” Satisfactory.” Each year, countless employees are evaluated on the job. Performance appraisals (PAs) occur with varying formality and structure and serve many administrative and developmental purposes within organizations. At a minimum, a PA is the evaluation and formal documentation of an individual’s job performance. This is often paired with a “feedback interview,” in which the rater (e.g., supervisor) discusses the evaluation with the employee. PAs are frequently used to support employment-related decisions about training, bonuses, raises, promotions, and terminations.

Due to their prevalence in organizations, critical value as documentation of employee performance, and role in employment decisions, PAs are often referenced in employment litigation. In most of these cases, an employee is the plaintiff bringing forward a discrimination claim against an organization. Over the past 40 years, there have been a number of quantitative and qualitative reviews of the presence of PA in case law (e.g., Feild & Holley, 1982; Malos, 1998; Malos, 2005; Werner & Bolino, 1997). The focus of those reviews was usually to identify relationships between characteristics of PA systems and court decisions. The majority of these examinations occurred in the 1980s and 1990s following changes in the legal protections afforded to employees (e.g., Title VII of the Civil Rights Act of 1964 [Title VII; CRA]; The Age Discrimination in Employment Act of 1967 [ADEA]; Pregnancy Discrimination Act of 1978 [PDA]).

The purpose of the present study is to provide an updated review of the role of PA in employment litigation. The present influence PA content and implementation characteristics have on court decisions will be revealed by examining five years (2014-2018) of case law. This

review extends previous reviews by examining more PA characteristics than previously considered (e.g., the position of the raters, rater training, the performance reflected in the ratings) and characteristics of the case itself (e.g., the inclusion of retaliation claims related to claims of discrimination). In addition, five previously untested hypotheses will also be examined.

Since reviews of case law are time-bound and the last quantitative and qualitative reviews were published in 1981 and 2005, respectively, this review provides a much-needed update and extension. The United States (U.S.) judicial system relies on legal precedent, and these decisions can be overturned or modified by later court decisions. This means that whereas the language of the law itself does not change; interpretation of the law may change. Further, examining more recent cases may reveal new ways that PAs are used as evidence in litigation.

In addition to legal considerations, there is also a growing movement for organizations to eliminate the use of PAs. Politics, biases, negative feedback, and logistical constraints, among other things, make the evaluation process challenging and undesirable from both the organization and employee perspective. As a result, popular press authors have advocated for abolishing PA (Coens & Jenkins, 2000), numerous companies claim to have done away with it (Buckingham & Goodall, 2015), and industrial-organizational (I-O) psychologists and human resource management (HRM) scholars have debated the pros and cons of doing so (Adler et al., 2016; Cleveland & Murphy, 2016; Hettal & Garza, 2016; Schleicher et al., 2018). Despite the recent trend to question the value of PA (Rock et al., 2014; Rock & Jones, 2016), the effect of eliminating PA remains unknown. This review provides some initial data on the extent to which organizations, employees, and judges rely on PA in employment litigation, as well as the extent to which organizational recommendations for legally defensible PA content and implementation appear in the case law.

2. PERFORMANCE APPRAISAL AND LEGAL CONSIDERATIONS

The practice of appraising the performance of employees has received much attention by I-O psychologists, HRM scholars, and practitioners alike. PA forms vary in formality and structure, but in most workplaces, PA is a formal, organization-initiated process that occurs annually or biannually (DeNisi & Pritchard, 2006). In practice, supervisors rate their subordinate's performance across a range of job-related criteria ideally identified through a job analysis. PAs are often a component of a larger performance management (PM) system (e.g., Levy et al., 2017; Schleicher et al., 2018). PM systems are the "continuous process of identifying, measuring, and developing the performance of individuals and teams" in order to facilitate the ability for performance to meet the organization's larger goals (Aguinis, 2013, p. 2). Although PA is only one piece of a larger PM system, it is clear that the measurement and evaluation of employee performance is a critical component of HRM.

2.1. Purpose of Performance Appraisals

PAs serve many administrative and developmental purposes within organizations. A survey of I-O psychologists employed in the private sector revealed that organizations use PAs for four primary purposes: (a) between-individual comparisons, (b) within-individual comparisons, (c) identification of developmental needs, and (d) documentation to support administrative decisions (Cleveland et al., 1989). Supervisors conduct between-individual comparisons in order to gauge employee performance relative to their peers. Within-individual comparisons reveal individual employees' strengths and weaknesses in order to identify training needs or promotion potential and to track changes in performance over time. Finally, PAs are used to document performance as well as defend personnel decisions (e.g., promotions).

Cleveland et al. (1989) found the four purposes were highly correlated with one another (r = .42-.67) suggesting many organizations use PA for more than one purpose. In fact, most organizations in that study reported using PAs for multiple purposes rather than one single organizational need.

In 2013, Aguinis added two more PA purposes to the list: strategic and organizational maintenance. The strategic purpose of PA is to link the organization's goals with the individuals' goals in a PM system. PAs reinforce behaviors aligned with organizational goals and correct actions that do not. PA criteria communicate desired behaviors and the organization's values to new employees, facilitating their onboarding to the organization. From an organizational maintenance and workforce planning standpoint, HRM use PA information to anticipate and respond to internal and external talent needs, assess training needs, and evaluate the effectiveness of HR interventions (e.g., training).

Over the years, there have been multiple empirical examinations focused on how the PA's purpose can influence the ratings or evaluations of an employee's performance (e.g., Boswell & Boudreau, 2001; Jawahar & Williams, 2006). In practice, appraisal feedback reveals how employees are performing in their role, areas where they can improve, and areas where they are excelling. Administratively, a positive evaluation may lead to a promotion or raise, whereas a negative evaluation may result in a performance improvement plan (PIP), a demotion, or termination. PAs might even be considered high-stakes' evaluations for many employees, because of the consequences associated with the ratings (e.g., termination). Given the importance of the decisions associated with PA to employees, researchers have proposed that within-individual ratings obtained for administrative decisions are more lenient than ratings for developmental use (Taylor & Wherry, 1951). Meta-analytic results have substantiated this claim

revealing that ratings gathered for administrative purposes are on average one-third of a standard deviation higher than ratings gathered for developmental purposes (Jawahar & Williams, 1997). It is likely that raters are intentionally more lenient in order to avoid the negative administrative consequences associated with ratings that are more accurate but potentially punitive (Ostroff, 1993).

Aguinis (2013) also noted that PAs are beneficial for supervisors, the organization, and employees. PAs benefit supervisors by clarifying expectations and standards for subordinates (Whitaker et al., 2007), improving communication with subordinates (Martin et al., 2015), and facilitating fair and job-relevant administrative decisions (Dipboye & de Pontibriand, 1981). When a PA is conducted appropriately (see Malos, 2005), the organization may be protected from lawsuits (Martin et al., 2000; Nickols, 2007), experience less misconduct (Werbel & Balkin, 2010), and organizational goals are likely to be clearer (Schleicher et al., 2018). Employees also benefit from role clarification (Whitaker et al., 2007), as well as increases in PA satisfaction (Keeping & Levy, 2000), motivation, self-esteem, self-insight, and performance (Kuvas, 2006).

Overall, PAs are an important tool for organizations and their use can have a direct impact on individual employees. Organizations often use PAs to fulfill multiple purposes simultaneously. Some of these needs (i.e., administrative decisions) result in significant changes and decisions for employees.

2.2. Legal Defensibility of Performance Appraisals

Performance appraisals serve as evidence of an employee's performance in court cases. As previously noted, one of the primary purposes of PAs is to be used as documentation and justification for administrative decisions including salary changes, promotion, and termination.

As such, supervisor ratings on annual PA forms are one of the primary sources of evidence or "proof" of how an employee performed on the job. Indeed, they are often the only tangible and formal evidence of performance available to both parties. Within a court case, either the organization (employer) or the employee, or both, can bring PA ratings forward as evidence to support their claim and/or defend decisions made.

When an organization is attempting to defend a personnel decision with a negative outcome for the employee (e.g., termination), they are likely to use PA ratings as evidence that an employee had unsatisfactory performance or a decline in performance. For example, organizations might use PA ratings to demonstrate that the employee's performance was the primary reason for termination or no promotion. Organizations might also use PA ratings of other employees to provide evidence that another employee was more deserving of a promotion or a distribution of ratings in an attempt to show how other employees were evaluated and how the focal employee fared relative to others.

Alternatively, or in addition, employees might use PA ratings as evidence of satisfactory performance to demonstrate they were deserving of a bonus, increase in pay, or promotion, or that a termination was not justified. Employees can also put forth the argument that unsatisfactory PA ratings are a form of retaliation, intentional discrimination or, despite appearing neutral, there was a systemically unfair impact on a protected class due to the PA's content or procedure (Malos, 2005).

2.2.1. *Previous Examinations of Case Law*

Over the last 40 years, there have been a number of reviews of case law involving performance ratings to identify the important determinants of court decisions. In a review of personnel practices Holly and Field (1976) noted the growing presence of PAs and PA ratings in

employment litigation following the CRA of 1964. First, qualitative examinations identified PA content and process issues (e.g., using PA as a basis for termination; the process of generating PA content) inherent in many PA systems (Basnight & Wolkinson, 1977; Cascio & Bernardin, 1981).

Subsequent quantitative examinations found courts decisions in favor of the organization were a function of several PA system characteristics (e.g., PA content based on a job analysis and written instructions provided to the raters; Feild & Holley, 1982; Feild & Thompson, 1984). Following these foundational studies, many reviews offered suggestions on how to create validated and psychometrically sound PAs (Barrett & Kernan, 1987; Kleiman & Durham, 1981; Wells, 1982). Several others noted the growing occurrence of PAs as evidence in employment litigation cases (Davidson, 1995; Martin & Bartol, 1991; Martin et al., 2000; Martin et al., 1986). Later publications echoed the previous findings using quantitative (Werner & Bolino, 1997) and qualitative (Malos, 1998) examinations of a larger number of cases. In many of the quantitative examinations, researchers conducted analyses of case decisions in which the decisions were regressed onto characteristics of the case (e.g., the PA was based on a job analysis) in order to determine which factors played a role in the judge's decision (e.g., Werner & Bolino, 1997). In the most recent qualitative review, Malos (2005) found organizations fared well when PAs incorporated job-related criteria and various aspects of due process and procedural justice.

In 1998, Malos synthesized the PA recommendations for organizations to date into two lists which are included in I-O psychology textbooks (Aguinis et al., 2018). Specifically, he identified six content recommendations and nine process recommendations based on multiple studies and previous proclamations of PA researchers (Ashe & McRae, 1985; Barrett & Kernan, 1987; Beck-Dudley & McEvoy, 1991; Bernardin et al., 1995; Burchett & De Meuse, 1985;

Cascio & Bernardin, 1981; Lubben et al., 1980; Martin & Bartol, 1991; Martin et al., 1986; Veglahn, 1993). The findings of previous reviews provide initial justification for hypotheses concerning PA characteristics that are legally defensible and thus will be reviewed in turn. Table 1 and Table 2 summarize the extant quantitative and qualitative literature respectively. Table 3 presents contextual details about the number and source of cases included in each examination

Conceptually, favoring the organization means the practice is expected to positively contribute to an organization's ability to defend corresponding personnel decisions. Functionally, when an employee pursues a claim in court, the decision can favor the employee (i.e., the court has supported the claim) or the employer (i.e., the claim is not supported). Importantly, a decision is made for each claim in a case.

Table 1

Prior Quantitative Examinations of Performance Appraisal Characteristics in Litigation

Recommendation	Citation for Study			
	Feild & Holley (1982)	Feild & Thompson (1984)	Miller, Caspin & Schuster (1990)	Werner & Bolino (1997)
PA Criteria should be…				
1. objective rather than subjective				
2. job-related or based on a job analysis	Supported	Supported	Insufficient Evidence	Supported
3. focused on behaviors rather than traits	Supported	Supported	Insufficient Evidence	Not Supported
4. within the control of the ratee				
5. specific functions, not global assessments	Supported	Supported	Insufficient Evidence	Not Supported
6. communicated to the employee				

Table 1 Continued

Recommendation	Citation for Study			
	Feild & Holley (1982)	Feild & Thompson (1984)	Miller, Caspin & Schuster (1990)	Werner & Bolino (1997)
The PA process should…				
1. be standardized and uniform for all employees within a job group				
2. be formally communicated to employees				
3. use multiple, diverse, and unbiased raters	Not Supported			Supported
4. provide written instructions and training to raters	Supported	Supported	Insufficient Evidence	Supported
5. contain thorough and consistent documentation across raters that include specific examples of performance based on personal knowledge	Not Supported			
6. provide employees notice of performance deficiencies and opportunities to correct them				
7. provide access for employees to review appraisal results	Supported	Supported	Insufficient Evidence	Supported
8. provide formal appeal mechanisms that allow for employee input				
9. establish a system to detect potentially discriminatory effects or abuses of the system overall				

Note. Descriptions of performance appraisal (PA) characteristics were standardized across studies. Empty cells indicate a recommendation was not considered. Supported indicates the recommendation was significantly related to decisions favoring organizations, not supported indicates the recommendation was not significantly related to decisions favoring organizations, and insufficient evidence reflects when information about the recommendation was recorded but there was insufficient evidence for analysis.

Table 2

Prior Qualitative Examinations of Performance Appraisal Characteristics in Litigation

	Recommendation	Citation for Study				
		Cascio & Bernadin (1981)	Kleiman & Durham (1981)	Barrett & Kernan (1987)	Dudley & McEvoy (1991)	Malos (1998) & Malos (2005)
PA Criteria should be…						
1.	objective rather than subjective	Supported			Not Supported	Supported
2.	job-related or based on a job analysis	Supported		Supported	Supported	Supported
3.	focused on behaviors rather than traits	Supported	Supported			Supported
4.	within the control of the ratee	Supported		Supported		Supported
5.	specific functions, not global assessments	Supported				Supported
6.	communicated to the employee	Supported	Supported			Supported
The PA process should…						
1.	be standardized and uniform for all employees within a job group			Supported		Supported
2.	be formally communicated to employees					Supported
3.	use multiple, diverse, and unbiased raters				Not Supported	Supported

Table 2 Continued

	Recommendation	Citation for Study				
		Cascio & Bernadin (1981)	Kleiman & Durham (1981)	Barrett & Kernan (1987)	Dudley & McEvoy (1991)	Malos (1998) & Malos (2005)
	The PA process should…					
4.	provide written instructions and training to raters		Supported		Not Supported	Supported
5.	contain thorough and consistent documentation across raters that include specific examples of performance based on personal knowledge			Supported		Supported
6.	provide employees notice of performance deficiencies and opportunities to correct them			Supported		Supported
7.	provide access for employees to review appraisal results		Supported		Supported	Supported
8.	provide formal appeal mechanisms that allow for employee input	Supported		Supported	Supported	Supported
9.	establish a system to detect potentially discriminatory effects or abuses of the system overall					Supported

Note. Descriptions of performance appraisal (PA) characteristics were standardized across studies. Empty cells indicate a recommendation was not considered. Supported indicates the recommendation was significantly related to decisions favoring organizations. Not supported indicates the recommendation was not significantly related to decisions favoring organizations.

Table 3

Cases and Recommendations from Previous Examinations of Performance Appraisal in Case Law

Citation for study	Year Range	No. of cases	Court Level (*N*)	Claims Examined (*N*)	Claim Decision (*N*)	Case Specifications	Additional Characteristics Related to Decisions for Organizations[a]
Cascio & Bernadin (1981)	Not Stated	Not Stated	Appeals Supreme	Title VII	Not Stated	Focus was on illustrative cases where one or more recommendations were violated	Appraisal instrument is validated and psychometrically sound
Kleiman & Durham (1981)	1965-1980	23	Not Stated	Title VII	Not Stated	Focus was on cases involving promotion decisions	Appraisal instrument is validated and psychometrically sound
Feild & Holley (1982)	1965-1980	66	State (2) District (46) Appeals (16) Supreme (2)	Title VII Race (40) Sex (15) ADEA (11)	Employee (31) Organization (35)	PA was used as the basis of making a personnel decision	Organization was nonindustrial Appraisal instrument is validated and psychometrically sound[a]
Feild & Thompson (1984)	1980-1983	31	District (21) Appeals (10)	Title VII Race (23) Sex (8)	Employee (9) Organization (22)	PA was used as the basis of making a personnel decision	Organization was nonindustrial[a]

Table 3 Continued

Citation for study	**Year Range**	**No. of cases**	**Court Level (*N*)**	**Claims Examined (*N*)**	**Claim Decision (*N*)**	**Case Specifications**	**Additional Characteristics Related to Decisions for Organizations**
Barrett & Kernan (1987)	1973-1986	51	Not Stated	Not Stated	Employee (10) Organization (41)	Focus was on cases involving termination decisions	
Miller, Caspin & Schuster (1990)	1968-1986	53	District (39) Appeals (14)	ADEA	Employee (12) Organization (25)	Focus was on ADEA cases	Employee was younger (40-49 years old)
Dudley & McEvoy (1991)	1980-1990	46	District (23) Appeals (21) Supreme (2)	Title VII Race (26) Sex (10) Race/Sex (4) Nat. Origin/ Race/Sex (1) Race/ADEA (1) ADEA (4)	Employee (22) Organization (21) Unknown (3)		No evidence of overt discriminatory behavior exists
Werner & Bolino (1997)	1980-1995	295	Appeals	Title VII Race (102) Sex (50) Combined (33) ADEA (109)	Employee (122) Organization (173)		PA procedures permitted the calculation of agreement across multiple raters

Table 3 Continued

Citation for study	Year Range	No. of cases	Court Level (*N*)	Claims Examined (*N*)	Claim Decision (*N*)	Case Specifications	Additional Characteristics Related to Decisions for Organizations
Malos (1998)	Not Stated	500+	Not Stated	Title VII CRA 1991 ADEA ADA Equal Pay Act	Not Stated	Focus was on illustrative cases	
Malos (2005)	2000-2005	Not Stated	Appeals	Title VII CRA 1991 ADEA ADA Equal Pay Act	Not Stated	Focus was on illustrative cases	

Note. Information about the case search and composition are presented as they were provided in the study. Depending on the circumstances a claim can be pursued in state court (state) or federal court. There are three levels of U.S. Federal Courts; District = U.S. District Courts, Appeals = U.S. Circuit Courts of Appeals, and Supreme = U.S. Supreme Court. Title VII = Title VII of the Civil Rights Act

2.2.2. ***Performance Appraisal Content***

Objective. Malos's (1998) first recommendation for legally sound appraisal criteria is to be objective rather than subjective. Most PAs are considered subjective due to their reliance on judgments made by raters. Nevertheless, the information rated (criteria) can vary in objectivity. Some more objective performance criteria include behavioral data such as sales volume, number of errors, accidents, as well as withdrawal behaviors such as lateness and absences. More objective criteria are preferred as they are measured more reliably, reducing the potential for criterion contamination. However, a major issue with more objective data is they are not exclusively a function of the employees' behavior and therefore are not fully under the control of employees (which directly contradicts the fourth content recommendation) and may suffer from criterion deficiency. A hypothesis aligned with this recommendation was not proposed due to the inherent subjectivity of the evaluations made when completing a PA.

Job-Related. Malos's (1998) second recommendation is for the PA to be job-related or based on a job analysis. Job analyses are the foundational bedrock of I-O psychology and a critical antecedent to the development of a valid PA system (Latham & Wexley, 1980). A job analysis is the collection of information about the work performed within an organization and the requirements needed to perform that work. In a job analysis, data are collected from subject matter experts to delineate the knowledge, skills, abilities, and other characteristics necessary to perform a given job (Cascio & Aguinis, 2010). In addition, each job is described in terms of three to five major duties and hundreds of tasks within each duty. A PA based on a job analysis is likely to incorporate behaviors necessary to complete tasks/duties identified in the job analysis. Correspondingly, PAs based on a job analysis are more likely to use appraisal criteria based on behaviors that are within the control of the ratee and related to specific duties on the job. Also,

these PAs are expected to have higher levels of face and content validity as well as less criterion deficiency.

In the first quantitative examination of PA system characteristics, Feild and Holley (1982) reviewed 66 discrimination cases. Judges were more likely to find in favor of the organization if a job analysis was used when developing the appraisal system. This finding has been replicated in several subsequent examinations (e.g., Feild & Thompson, 1984; Werner & Bolino, 1997). Correspondingly, due to the historical significance of job analysis and its usefulness for demonstrating the job-relatedness of a PA, when the PA is based on a job analysis, the organization is more likely to successfully defend its actions.

Hypothesis 1: When the appraisal content is based on a job analysis, litigation decisions will be significantly more likely to favor the organization than the employee

Specific and Controllable Behaviors. The next three PA recommendations also concern the content of the PA form: measuring behaviors over traits, using criteria within the control of the employee, and evaluating specific functions over global functions. These three recommendations are closely aligned and can be best addressed by basing PA content on a job analysis. Early examinations found that organizations who used behavioral criteria rather than traits were more likely to have favorable decisions (Feild & Holley, 1982; Feild & Thompson, 1984). Miller, Caspin and Schuster (1990) were unable to find a sufficient amount of job analysis information for analyses, and Werner and Bolino (1997) were unable to replicate this finding. Nonetheless, it is clear that behaviors (e.g., showing up on time) are preferred over traits (e.g., conscientiousness), as they are within the control of the employee. The use of specific job functions also reflects an alignment with the employees' job descriptions. These recommendations have remained a constant among qualitative reviews (e.g., Malos, 2005).

Correspondingly, when PA content contains specific, controllable behaviors, judges may be more likely to find in favor of the organization.

Hypothesis 2: When appraisal content is (a) specific rather than vague, (b) controllable vs. uncontrollable, and (c) behaviors rather than traits, litigation decisions will be significantly more likely to favor the organization than the employee.

Communicated to the Employee. Malos's (1998) final recommendation for the content of a PA is that the criteria should be communicated to the employee. Ideally, criteria are conveyed to the employee in advance, well before they are evaluated. This allows for employees to align their behaviors with what is valued and understand the standards that they must meet (Cleveland et al., 1989). When organizations communicate PA criteria to employees in advance, judges may be more likely to find in their favor.

Hypothesis 3: When appraisal content is communicated to the employee, litigation decisions will be significantly more likely to favor the organization than the employee.

2.2.3. ***Performance Appraisal Process***

As noted earlier, Malos (1998) also provided nine procedural recommendations or suggestions on how to implement PA. In 2005, Malos performed an updated examination of cases and concluded that these same recommendations remained essential to a legally defensible system. All of these procedural recommendations can be related to the concept of justice. Justice and perceptions of fairness are essential to the effective functioning of organizations and the satisfaction of their employees (Greenberg, 1990; Greenberg & Colquitt, 2005). Justice has been conceptualized as having three components: the fairness of the outcome achieved (distributive justice), the fairness of the process used to achieve the outcome (procedural justice), and the way management behaves towards the recipient of justice (interactional justice; Cohen-Charash &

Spector, 2001). While distributive justice and interactional justice are important to employee outcomes, procedural justice most closely underlines Malos' recommendations. Indeed, early studies of justice found that satisfaction with fictitious employment case decisions was independently influenced by the individual's perception of the fairness of the process used to make the judgment (Thibaut & Walker, 1975).

The fairness of the process by which organizational outcomes are achieved is often the most important determinant of total perceived organizational justice (Lind & Tyler, 1988). As such, the reactions to procedural decisions are more organization-focused rather than outcome-focused and are directed at the whole organization resulting in reactions that affect outcomes such as organizational commitment and job satisfaction (Cohen-Charash & Spector, 2001). According to Leventhal (1980), there are six rules which result in procedures that are considered fairer: (1) consistent allocation across people and time; (2) prevention of personal bias in decision making; (3) use of accurate information; (4) the opportunity for corrections; (5) the equal assessment of needs, values, and outlooks of all parties; and (6) the compliance with moral and ethical values of the perceiver. These six rules relate directly to many of Malos' procedural recommendations. The perception of procedural justice is important as employees who feel they have been mistreated are more likely to file an employment discrimination claim (Goldman, 2001). Three of Malos's procedural recommendations (i.e., the use of multiple raters; written instruction for raters; the ability to review ratings) have been previously examined in quantitative analyses (see Table 1). Based on those previous reviews and the value of procedural justice, organizations that follow these procedural recommendations are more likely to receive favorable litigation decisions.

Procedural Standardization. The first PA procedural recommendation offered by Malos (1998) is for the PA procedures to be standardized and uniform for all employees within a job group. Evaluating employees on the same criteria facilitates fair between-person comparisons and is more practical than tailored evaluations for each employee. This recommendation is likely to facilitate correspondence with Leventhal's (1980) rule of consistent allocations across people and time. Likewise, in a recent examination of 312 court cases concerning selection practices, Williams, Schaffer, and Ellis (2013) concluded that the legal landscape required organizations to adhere to fair and consistent procedures.

Hypothesis 4: When the PA is standardized and uniform for all employees within a job group, litigation decisions will be significantly more likely to favor the organization than the employee.

Formal Procedure Communication. Malos's (1998) second PA procedural recommendation is to formally communicate the organization's PA procedure to employees. If the PA procedure is not formally communicated to employees, there is a greater probability that it will not be implemented consistently across employees within the organization, thus administration will not be standardized (the first procedural recommendation). Further, the evaluation and its implementation may come as an unpleasant surprise to employees.

Hypothesis 5: When appraisal procedures are formally communicated to employees, litigation decisions will be significantly more likely to favor the organization than the employee.

Multiple, Diverse, Unbiased Raters. PA ratings require the raters to observe employee behavior and make judgments. Observation includes the detection, perception, and recognition of performance events, as well as recall of the observations. Judgment includes the cognitive

organization of observed events into performance categories, evaluation of those events, and integration of the events into performance ratings (Thornton & Zorich, 1980).

Procedurally, observations, judgments, and the ratings can be generated by several sources. In most organizations, the immediate supervisor is responsible for observing employee behaviors and making PA ratings. Supervisors are usually the best source to evaluate an employee's contribution to organizational goals and generally have higher correlations with objective measures of performance than other rating sources (Becker & Klimoski, 1989).

Malos (1998) recommends that PAs should include information from multiple, diverse, and unbiased raters. Thibaut and Walker (1978) also suggested that using multisource ratings could result in a more legally defensible PA process. They proposed that PAs informed by the evaluations of multiple raters could provide a corroborative defense of the organization's administrative decisions. After all, more raters results in more ratings and therefore more data on which to make decisions. Multiple raters allow for the computation of interrater agreement, interrater reliability, as well as differences between the raters which may prompt revisions to ratings. Further, it has also been proposed that the accuracy and perceived usefulness of PA is positively associated with rater agreement (Meyer, 1980).

Previous case law reviews have investigated the use of multiple PA raters in litigation. Feild and Holley (1982) did not find a significant relationship between multiple raters and case decisions. This may have been due to the small number of cases where multiple raters were used. Later, Werner and Bolino (1997) found that cases were more likely to be found in favor of the organization when multiple raters agreed on the ratings. Subsequent to these reviews, starting in the late 1990s, many organizations adopted systems like 360-degree appraisals that encompass multiple raters from different sources familiar with the employee and their performance (Rynes

et al., 2005). Given the increased practice of multiple raters and the advantages associated with using multiple raters, organizations are expected to benefit from this practice.

Hypothesis 6: When employees are reviewed by multiple raters, litigation decisions will be significantly more likely to favor the organization than the employee.

Written Instructions and Rater Training. Training raters has often been advised as a method to reduce the bias and error in ratings. Previous examinations of rater training in case law have focused primarily on the provision of written instructions to raters. Written instructions on how to complete appraisals can assist with ensuring that PA procedures are perceived as fair. These written instructions typically include information about the timeline of PA procedures, the format of the feedback, and other explicit instructions about how the PA should be completed. By providing explicit instructions, the generation of PA ratings is more likely to be standardized and uniform. Feild and Holley (1982) found judges were more likely to find in favor of the organization if evaluators were given written instructions on how to complete appraisals. This finding has been replicated in several subsequent examinations (Feild & Thompson, 1984; Malos, 1998; Werner & Bolino, 1997).

Over the years, various forms of rater training have been developed. Whereas written instructions for raters described in previous examinations included only basic procedural information, rater training encompasses a range of different approaches. Some rater training focuses on informing raters of errors (e.g., halo effect, leniency bias, and central tendency) that may occur during the PA process so that they will avoid these practices when appraising performance (i.e., rater error training). Other training methods are designed to teach managers the meaning of each performance dimension. Some rater training includes formal "calibration" sessions which encourage managers to meet together to compare the performance and

corresponding ratings of their employees in order to reach consensus (DeNisi & Pritchard, 2006). Overall, frame-of-reference training has proven to be the most effective method for improving rating accuracy (Roch et al., 2011; Woehr & Huffcutt, 1994). Newer training methods address personal biases that raters may have towards members of different races or sex (e.g., unconscious bias training). Training to reduce bias may increase perceptions of procedural justice by reducing personal biases that can enter into decision making (Leventhal, 1980). Given the time between previous case law examinations and developments in rater training, there has yet to be a systematic evaluation of rater training on PA ratings in case law. Correspondingly, the presence of written instructions and rater training are proposed as significant determinants of case decisions.

Hypothesis 7: When specific written instructions or training is provided to raters, litigation decisions will be significantly more likely to favor the organization than the employee.

Performance Documentation and Correction Opportunities. Malos's (1998) next procedural recommendations are that a PA should include thorough and consistent documentation and provide employees with the opportunity to correct deficient performance. This documentation should include specific instances of performance or behaviors based on personal knowledge or observation. This way the raters are also more likely to provide feedback that is based on behaviors rather than traits. Relying on documentation of performance is likely to be perceived as fairer than relying on memory when generating ratings. Procedures that entail the thorough documentation of behaviors across several PAs to identify performance deficiencies are likely to be perceived as more procedurally just (Leventhal, 1980). Correction opportunities may also vary in formality from suggestions for improvement during a PA feedback interview to formal PIPs with clearly outlined performance expectations and set deadlines. Correspondingly,

Barrett and Kernan (1987) previously found that litigation decisions were more likely to favor organizations when employees were provided notice of deficiencies and opportunities for correction. Therefore, it is hypothesized that thorough documentation of performance and offering corrective opportunities will result in more favorable decisions for the organization.

Hypothesis 8: When organizational PA procedures require thorough and consistent performance documentation, the litigation decisions will be significantly more likely to favor the organization than the employee.

Hypothesis 9: When organizational PA procedures provide employees with an opportunity to correct these deficiencies, litigation decisions will be significantly more likely to favor the organization than the employee.

Review Appraisal Results. As an extension of the previous procedural recommendations, Malos also recommended that employees be given access to review the appraisal results (Malos, 1998). In other words, employees should be given the opportunity to see their ratings and comments. Oftentimes, managers discuss the ratings with the employee and provide additional information during feedback interviews. Ideally, this discussion presents the employee an opportunity to ask questions and contribute to the conversation. Early examinations of appraisal systems found that the effectiveness of the PA is positively related to the level of employee involvement (Burke & Wilcox, 1969; Meyer & Walker, 1961). Participation in the appraisal process also results in increased appraisal system acceptance, perceptions of appraisal fairness, overall appraisal satisfaction, perceived support, and acceptance of negative feedback (Cawley et al., 1998; Roberts, 2003). When employees are given the opportunity to review the results of their appraisal, supervisors can also be certain that employees are aware of how their performance was appraised.

Previous reviews (e.g., Feild & Thompson, 1984) have revealed that judges expect organizations to give employees the opportunity to review the results of their PA ratings. In cases where organizations permitted employees to review their ratings, judges were more likely to find in favor of the organization (Feild & Thompson, 1984; Malos, 1998; Werner & Bolino, 1997). Feild and Thompson (1984) concluded that the PA process was viewed more favorably for the organization when the appraisal results were reviewed because it provided an opportunity for the employee and manager to increase transparency and fairness. Malos's (1998, 2005) subsequent qualitative reviews also identified the opportunity to review ratings as being a critical procedural justice element. Therefore, the ability to review PA results remains a relevant consideration.

Hypothesis 10: When employees are allowed to review their ratings, litigation decisions will be significantly more likely to favor the organization than the employee.

Formal Appeal and Abuse Mechanisms. The final PA procedural recommendation offered by Malos (1998) is to provide employees with a formal appeal mechanism that allows for employee input. If an employee has reviewed the results of their PA and believes that they were evaluated unfairly, then employees should have the opportunity to refute evaluations they believe are inaccurate. This appeal can be done directly to their rater or to an external source if the employee believes their complaint will not be addressed adequately by the rater. A dynamic PA feedback interview may present an opportunity for employees, raters, and management to review the PA given and discuss areas of disagreement. Sometimes the decisions made after the PA review can result in a modified PA rating or decision. The opportunity to refute an evaluation that an employee believes is unfair will likely increase the employee's perceived procedural justice. Additionally, court decisions may consider this procedural element an internal resolution

to employment disputes related to PAs and be more likely to favor organizations that allow appeals and investigate potential abuse.

A system that detects potentially discriminatory effects or abuses of the system overall will also likely increase the employee's perceived procedural justice. These systems could audit the organization's PA practices by examining PAs across the organization. For example, the audit could examine whether the process has a disproportionate impact on any groups. Audits could also reveal whether a particular manager is more prone to rating errors. If rater agreement is low, audits could also prompt meetings to encourage rating consensus. Systems to detect and reduce bias may increase perceptions of procedural justice as employees may see such a system as preventing bias in decision making.

Hypothesis 11: When organizational PA procedures (a) allow employees to appeal PA decisions, and/or (b) there is evidence that an employee pursued an appeal, the litigation decisions will be significantly more likely to favor the organization than the employee.

Frequent Feedback. Frequent appraisals are often recommended so that the PA ratings can provide timely and developmentally-pertinent information to employees. Criticism of the traditional annual PA comes, in part, from is lack of timely feedback. It is recommended that PAs should be scheduled to align with the completion of major projects or occur continually in shorter intervals that are appropriately matched to the job performed (Aguinis et al., 2012).

Instead of an annual meeting, more frequent feedback gives employees more opportunities to improve performance. Some research supports this and suggests that more frequent and continuous feedback improves task performance and the ability for employees to learn from the feedback (Salmoni et al., 1984). Others have found that frequent feedback can overwhelm an individual's cognitive resources and reduce task effort (Lam et al., 2011).

Despite the lack of empirical evidence supporting appraisals occurring more frequently than annually, the relationship between the frequency of evaluations and case decisions has been identified as an important consideration. While, contrary to their expectations, Werner and Bolino (1997) found that appraisal frequency was unrelated to the case decisions, they still emphasized frequent PAs as an important practice. Many others also suggest the implementation of more frequent PA as an improvement to the PA process (Kacmar et al., 2003; Norris-Watts & Levy, 2004; Steelman et al., 2004). Thus, this advice may have increased the frequency of appraisals in the time that has elapsed since the previous examination.

Hypothesis 12: The frequency of evaluations will be positively related to litigation decisions favoring the organization.

Adherence to Recommendations. Many of Malos's (1998) content and procedural recommendations are related to one another. If the content of a PA is not valid, then the value of the PA is limited. If the procedures are unfair, then the utility of the PA to correct and improve performance is equally limited. Many procedural recommendations follow logically (i.e., in order to appeal one or more PA rating, an employee must be allowed to review the results of the PA). Thus, organizational adherence to more of these content and procedural recommendations should be optimal.

Hypothesis 13: When there is evidence that an organization followed a greater number of Malos's content and procedural recommendations, the litigation decisions will be significantly more likely to favor the organization than the employee.

.

3. PERFORMANCE APPRAISAL IN LITIGATION

As stated earlier, employees who feel they have been mistreated are more likely to file an employment discrimination claim (Goldman, 2001). Most cases that utilize PAs as evidence involve a claim of discrimination and most previous examinations of the utilization of PA within litigation have focused on discrimination claims (e.g., Barrett & Kernan, 1987; Basnight & Wolkinson, 1977; Cascio & Bernardin, 1981). This, is due, in part, to the large number of discrimination claims that are pursued each year. In 2019 alone, 72,675 charges of discriminatory behavior were filed with the U.S. Equal Employment Opportunity Commission (EEOC; EEOC, n.d.a).

PA ratings are essentially judgments by raters and they are subject to the anchoring and adjustment heuristic (Tversky & Kahneman, 1974). According to this heuristic, when making a judgment an individual begins with a set reference point (i.e., an anchor) and by gathering additional information, the individual makes incremental adjustments before arriving at their final assessment. For this reason, judgments of employee performance are subject to the same process and it is understandable that ratings often suffer from persistent rater errors and biases (Huber et al., 1987; Saavedra & Kwun, 1993; Thorsteinson et al., 2008). Ratings may be influenced by rater bias or stereotypes. Indeed, examinations of ratee sex and race have revealed that demographic characteristics of the ratee have small influences on the ratings assigned to them (Pulakos et al., 1989). The race of the rater has also been shown to influence ratings with Black and White raters giving significantly higher ratings to members of their own race (Kraiger & Ford, 1985; Stauffer & Buckley, 2005). Undoubtedly, the influence of bias and stereotypes can result in potentially unlawful discrimination.

Discrimination is the unjust or prejudicial treatment of individuals based on membership in protected classes. There are two recognized types of illegal employment discrimination. *Disparate treatment* claims arise when an employee or applicant asserts that they were treated differently in a substantive way due to their membership within a protected class compared to other similarly qualified or performing individuals'. To do so, the applicant or employee must show that the difference in treatment was due to a discriminatory motive. The other recognized type of discrimination is disparate impact. *Disparate impact* is discrimination that results from employment practices that, while seemingly neutral on their face, have a discriminatory effect on some protected class. Disparate impact cases require the applicant/employee demonstrate that an employment practice had a negative effect on them, due to their membership to a protected class.

3.1. Legal Discrimination Protections

The Civil Rights Act of 1866, now referred to as Section 1981, provides federal protection against race discrimination in all contracts, including employment contracts. Section 1981 is only applicable to claims of race-based disparate treatment and not disparate impact. A Supreme Court decision in 2008 also extended protection to individuals experiencing retaliation after making a Section 1981 claim (CBOCS West, Inc. *v*. Humphries, 2008). Legally, Section 1981 applies to all private organizations, state, and local governments regardless of their size Section 1981 can also offer legal protections to individuals, like contractors, who would not be legally considered employees.

Title VII established by the CRA of 1964 (Title VII) and amended in 1991, prohibits discrimination against applicants and employees based on their race, color, religion, sex, and national origin. The CRA was extended to include the PDA (1978) which prohibits

discrimination based on pregnancy, childbirth, and related medical conditions. In 1968, the ADEA extended protection to those 40 and older. In 1990 (and amended in 2008), the Americans with Disabilities Act (ADA) prohibited discrimination on the basis of one or more disabilities. Employees may claim that some employment-related decision was actually an adverse employment action when their jobs, or ability to do their job, is affected negatively. Whether the employment-related decision is considered an actionable adverse employment action, or not, is a legal consideration. Decisions commonly protected under these laws include hiring, salary, benefits, promotion, training, and placement decisions. Adverse employment actions do not include minor inconveniences or alterations on the job.

Unlawful discrimination protections under these laws also include harassment based on membership to a protected class and retaliation against employees. Harassment is deemed unlawful when (1) enduring the offensive conduct becomes a condition of continued employment or (2) the conduct is severe or pervasive enough to create a work environment that a reasonable person would consider intimidating, hostile, or abusive. Employees must demonstrate that they felt their employment conditions were altered and that a reasonable person would objectively agree. The EEOC also considers claims of retaliation against employees who engage in a protected activity (e.g., asserted their rights to be free from employment discrimination). Employees who experience negative employment outcomes after asserting EEOC rights may put forward an additional claim of retaliation. EEOC protected rights can include participating in or filing an EEOC charge, communicating concerns of discrimination to management, and requesting accommodations.

It is important to note that, while many cases use the language "protected class" to refer to membership within protected groups, all U.S. citizens are subject to the same protections. Critically, this means it is illegal to discriminate against anyone (majority or minority members) by virtue of belonging to a protected class. Also, while an individual may make any number of discrimination, retaliation, or harassment claims, the legal procedure and burdens of proof differ. As this review is primarily focused on the content and process of PA, the U.S legal structure, litigation process, and burdens of proof are described only briefly.

3.2. Litigation Process

Litigation is the process of taking legal action by filing a charge to be heard in court. In litigation, the two primary parties are referred to as the plaintiff and the defendant. In most employment law cases, the employee is the plaintiff and the organization is the defendant. The plaintiff is the individual (or individuals) who files the lawsuit and attempts to take the case to court. The defendant is the party the plaintiff is raising one or more claims against. Plaintiffs seek legal remedy against the defendant through a court judgment (i.e., decisions) on the raised claims.

The United States has both federal and state court systems that operate at different levels of authority. Cases are first heard in a trial court in the state or federal system (i.e., the state trial courts and U.S. District Courts). At this lowest level, the judge or jury makes decisions about the plaintiff's claims. If the judge does not perceive there is sufficient grounds for a case, the judge may grant a motion from either the defendant or the plaintiff for summary judgment. A successful motion for a summary judgment renders a trial unnecessary. Whereas it is common for the plaintiff to bring forward multiple claims, a court may find that some claims have more

evidence than others. Thus, the court must evaluate each claim individually and rule in favor of the plaintiff or the defendant for each claim.

The decision made in the lower court can be appealed by referring the case to a higher court, a U.S. Federal Court of Appeal or a state appellate court. Cases at the appellate court level are cases that have reached a trial decision or have been subject to summary judgment. Either party can raise the case to a court of appeal based on an error in the trial procedure, error in interpretation of the law, or other grounds. The judges in the appellate court will review the evidence presented in the lower court and make a judgment about each contested claim decision based on the standard of review.

The standard of review prescribes the level of scrutiny and deference the appellate court can apply to the lower court decision. Under a "*de novo*" review, the appealing party is directing the appellate court to look at the issues of law from the lower court (i.e., review whether the trial court correctly applied the law) by reviewing the case anew. A "clearly erroneous" standard of review is used when the appellant believes that a fact (i.e., an event or circumstance) has been incorrectly considered by lower court and they must prove this to the appellate court.

The appellate court has higher authority and the decisions made in these cases are considered precedential. Appealed cases can result in several decisions. If the lower court's decision is *affirmed* then the original decision is not changed. When the decision is *reversed* the lower court's decision is overturned. The appellate court can also *remand* the case to back to the lower court. A remanded decision can include instructions for how the lower court should take further action (e.g., reconsider the facts of the case). A case can be remanded, reversed, and affirmed either in whole (all claims) or in part (some claims). In this way, the claims made

within a single case could all result in different decisions. After reaching a decision, the appellate courts may issue a written opinion. An opinion summarizes the facts of the case and the rationale behind the decisions of the court. The facts of an employment-related discrimination claim within a judge's opinion (case law narrative) can often include information relevant to PA.

3.3. Equal Employment Opportunity Commission

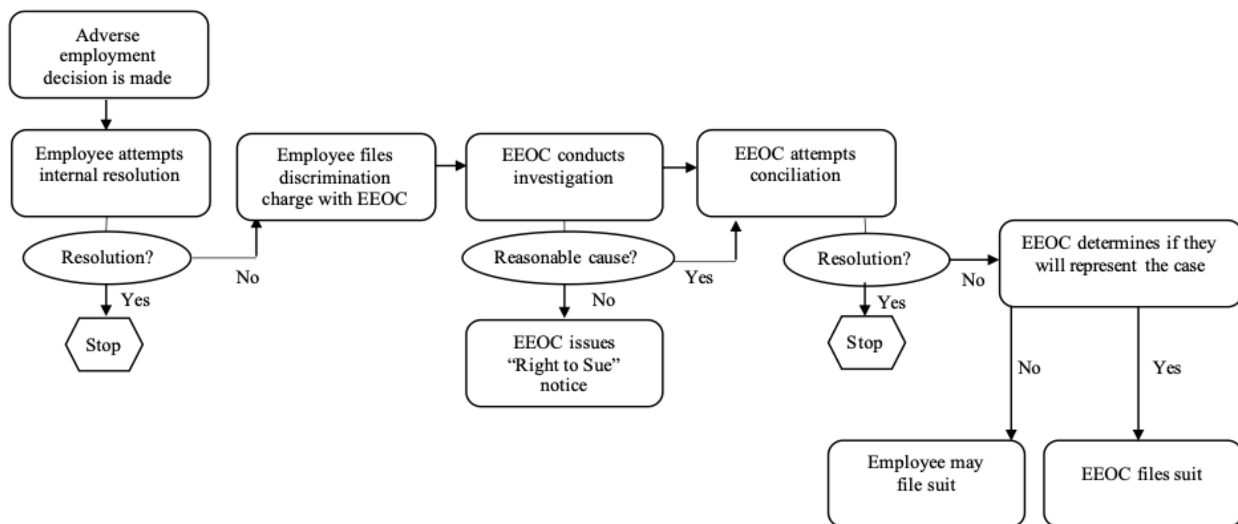
Most employees of organizations with at least 15 employees are covered by federal anti-discrimination laws (the exception being the requirement for organizations to have at least 20 employees for ADEA claims); if an employee believes that they have been discriminated against on the basis of their membership to one or more of the protected classes or conditions, they must file their complaint with the EEOC or a lawyer in a timely manner. Timely manner is defined by the EEOC and varies based on state and local agency laws. In general, an individual who believes s/he was discriminated against has 180 days to submit a complaint. (Federal employees and applicants must file complaints within 45 days.) If there are multiple discrimination claims or ongoing behavior like harassment, charges must be filed within 180 days of the last occurrence of each event. An eligible employee who has missed their deadline to submit a claim through the EEOC may qualify to make a claim through Section 1981. According to Title VII, retaliation against those who make claims is legally prohibited and is also subject to timely reporting policies.

The process the EEOC follows to review claims of discrimination is illustrated in Figure 1. After filing one or more discrimination claims, the EEOC will review the claim and request information from the accused organization. They will determine if (1) the laws regarding protected classes apply to the claims, (2) the event occurred within the required time frame, and

(3) there appears to be sufficient evidence of any violations of the law. The EEOC will also notify the organization of the charge and attempt to achieve conciliation between the organization and the individual. If mediation fails or the charge is not sent to conciliation, EEOC may pursue a lawsuit or they may advise the individual to pursue litigation independently. The EEOC brings claims against organizations only when they have sufficient evidence to support the case. If the EEOC deems there is insufficient evidence, individuals still have the right to sue without EEOC representation. In fact, most discrimination cases are filed by individuals or as class actions.

Figure 1

Simplified Equal Employment Opportunity Commission Process



Note. This figure illustrates a simplified EEOC process for examining discrimination charges and determining if they will represent the case and file suit.

Employees who are eligible to make Section 1981 claims of discrimination have several procedural advantages compared to Title VII claims. First, employees can immediately pursue litigation and do not have to report to the EEOC. Second, employees have a period of four years after experiencing the racially motivated discrimination to file a lawsuit. Finally, the monetary damages that can be awarded are not limited.

3.3.1. ***EEOC Represented Cases***

While the EEOC receives many claims of discrimination, as noted earlier, only a few cases are directly represented by the EEOC each year. When the EEOC investigates a charge and determines there is reasonable cause to believe that discrimination has occurred they first attempt conciliation. If conciliation between the two parties fails, the EEOC may decide to litigate and file a lawsuit against the organization. When deciding whether to file, the EEOC first considers the severity of the violation, the legal issues of the case, and the wider impact the lawsuit could have on future employment discrimination protections (EEOC, n.d.b). Of the 76,418 charges of discriminatory behavior filed with the EEOC in 2018, the EEOC only filed 199 lawsuits on behalf of 209 individuals and parties (EEOC, n.d.c). While the outcomes of these specific cases are still unknown, in 2018 the EEOC resolved 141 lawsuits and won approximately 95 percent of the district court cases (EEOC, n.d.c).

When the EEOC pursues litigation, the employee/plaintiff has the benefit of being represented by an entity experienced with enforcing discrimination protections. EEOC resources are used which minimizes the financial burden or strain the employee may have otherwise experienced. These benefits have proven to significantly increase the odds of a favorable litigation decision for the employee (Songer et al., 1999). Given the number of charges that the

EEOC investigates each year, the number of lawsuits filed, and their success rate, it would follow that EEOC lawyers are more likely to bring litigation against organizations when they believe they have a strong case and therefore would be more likely to win.

Hypothesis 14: *When the plaintiff is represented by the EEOC, litigation decisions will be significantly more likely to favor the employee than the organization.*

3.3.2. ***Retaliation***

The number of retaliation claims made to the EEOC is on the rise. Since 2008, retaliation claims are the most frequently alleged basis of discrimination (El Kharzazi et al., 2014). In the 2019 fiscal year, 39,110 retaliation charges were filed with the EEOC reflecting over 53 percent of all charges received by the EEOC (EEOC, n.d.a). Successful retaliation findings resulted in over 192 million dollars paid by organizations (EEOC, n.d.c). While a retaliation claim is dependent on an initial protected action (often a discrimination claim), the success of each claim is independent of the other (El Kharzazi et al., 2014).

When adverse employment actions occur to an otherwise satisfactorily performing employee after engaging in a protected activity, PA evidence could be used to demonstrate retaliation. PA evidence of changing performance scores may be instrumental to meeting the burden of a retaliation claim. Historically, a reduction in PA ratings over time is not considered significant enough to be an adverse employment action unless it is directly related to an outcome with greater impact (i.e., a poor PA that results in termination). Similarly, if the PA rating drops after engaging in a protected activity, and the PA is tied to some other significant employment outcome, this could also prove to be evidence of retaliation. Retaliation related to PAs

encompasses reduced ratings, pretext for a more severe decision, or even fabrication of PA rating data.

Hypothesis 15: *When there is evidence that performance appraisal ratings decreased or were altered after an employee filed a claim or action has been submitted, litigation decisions will be significantly more likely to favor the employee than the organization.*

3.4. Public and Private Sector Organizations

Employees in both the private and public sectors bring forth discrimination claims to the EEOC. Organizations in the public sector are government owned and operated. These public organizations are run and operated by federal, regional, state and local bodies to provide services to the public. Private sector organizations are privately owned. The distinctions between the two sectors are particularly important when considering the types of claims employees may bring forth. Employees in the public sector are granted additional rights that private sector employees are not afforded.

One critical difference is “at will” employment. In the private sector an at-will employee can be fired for any reason and organizations are not required to provide notice. Organizations in the public sector must have “just cause” and provide notice when disciplining or firing employees. Employees are also able to challenge personnel decisions. This “just cause” requirement should encourage managers to clearly communicate policies, expectations, and consequences before making employment decisions. Performance and behavioral issues should also be well-documented to defend in the event that an employee challenges the decision.

The role PA plays in private and public sector claims remains unexamined. Terpstra and Honorée (2016) found mixed results when examining the decisions of discrimination claims

made in the federal, local, and private sectors. While race claims made against public sector organizations were much less likely to succeed, the public sector fared less well in cases of sex and age discrimination. Despite these mixed results, it is hypothesized that the need to justify and document reasons behind personnel decisions will lead to more favorable decisions for public organizations.

Hypothesis 16: When the defendant is a public sector organization, litigation decisions will be significantly more likely to favor the organization than the employee.

3.5. Evidence of Satisfactory Performance

Authors of previous examinations have framed their hypotheses and analyses around decisions that favor the organization. As an extension of this, there are important yet unexamined performance-related variables that may reveal how PA can also be used by employees to substantiate their discrimination claims.

In cases where employees are aiming to prove that their negative employment outcome was unrelated to their performance, they may be more successful if they can demonstrate that their performance was satisfactory. PA can serve as a record of performance over time and ratings which are consistently high will reflect favorably on the employee. If there is no evidence to support a decrease in performance or a specific instigating event, an organization may have a more difficult time defending its negative personnel action. Thus, it is likely that when an employee presents a PA that documents satisfactory performance, the judge is more likely to rule in favor of the employee.

Hypothesis 17: *When there is PA evidence documenting satisfactory performance, litigation decisions will be significantly more likely to favor the employee than the organization.*

3.6. Examples of Performance Appraisals in Discrimination Cases

Since the focus of many of the discrimination claims do not directly involve the PA itself, providing and reviewing PA information is not the direct goal of the court or the case narratives. The amount of PA information, PA content and procedural elements, and information about the employee's performance is case-dependent. Three illustrative example cases are described next.

Godwin v. Wellstar Health Systems (2015) was an appeal of summary judgment against a 63-year-old employee who claimed she was terminated and retaliated against on the basis of her age and disability. In the process of reviewing the factual background (i.e., a "clearly erroneous" review of the trial case), evidence of one positive PA, nine negative PAs and many mixed PAs (totaling 23 evaluations) were presented as documentation of her performance in her 12-year tenure. This evidence was used by both the employee and WellStar. WellStar was able to successfully argue that the employee's performance continued to decline after she was given several opportunities to improve. The employee unsuccessfully argued that her supervisors had artificially lowered her ratings as pretext for her termination. The trial decision for summary judgement was affirmed.

Unlike the previous example, the presence of PA information can often be limited. In a *de novo* review, *Henry v. Federal Reserve Bank* (2015), an employee argued that he experienced a hostile work environment due to his religion and was retaliated against after he went to the EEOC with his charges. The only information included in the opinion that related to PAs was the mention of a mostly positive PA written by the supervisor which noted that the employee was "rigid". Ultimately, the trial decisions were affirmed and Henry failed to support his retaliation claim.

In another case, *Martin v. Eli Lilly & Co* (2017), the defendant appealed the district court's decision to deny summary judgment as a matter of law and the Appellate Court conducted a *de novo* review of a former employee's ADA claim. As a result of a company reorganization the employee's territory was assigned to another sales associate due to her higher "HR Index score". The employee argued that the three negative annual PAs constituted as an adverse employment action because the results of the PA were included among the criteria used to make the "HR Index Score". While a jury found in favor of Martin's ADA claim because the evaluations were adverse and they contributed to her displacement, the appellate court reversed the jury's decision. The appellate judges cited precedent which clarified that "Negative performance evaluations, standing alone, do not constitute adverse employment action" (*Lucas v. WW Grainer, Inc.,* 2001) and argued that since Eli Lilly & Co. offered the employee a lateral position and relocation expenses there was no adverse employment action. While the case that established the precedent that PAs alone cannot constitute an adverse employment action was concerning an ADA claim, the same precedent has been applied to Title VII claims.

As these three illustrative cases demonstrate, the case narratives contain varying amounts of PA information and detail. Nonetheless, as previous reviews have already illustrated, PA information is available in case narratives. A compilation of all of the hypotheses is presented in Table 4.

Table 4

Table of Hypotheses

No.	Hypotheses
Hypothesis 1:	When the appraisal content is based on a job analysis, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 2:	When appraisal content is (a) specific rather than vague, (b) controllable vs. uncontrollable, and (c) behaviors rather than traits, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 3:	When appraisal content is communicated to the employee, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 4:	When the PA is standardized and uniform for all employees within a job group, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 5:	When appraisal procedures are formally communicated to employees, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 6:	When employees are reviewed by multiple raters, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 7:	When specific written instructions or training is provided to raters, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 8:	When organizational PA procedures require thorough and consistent performance documentation, the litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 9:	When organizational PA procedures provide employees with an opportunity to correct these deficiencies, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 10:	When employees are allowed to review their ratings, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 11:	When organizational PA procedures (a) allow employees to appeal PA decisions, and/or (b) there is evidence that an employee pursued an appeal, the litigation decisions will be significantly more likely to favor the organization than the employee.

Table 4 Continued

No.	Hypotheses
Hypothesis 12:	The frequency of evaluations will be positively related to litigation decisions favoring the organization**.**
Hypothesis 13:	When there is evidence that an organization followed a greater number of Malos's content and procedural recommendations, the litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 14:	When the plaintiff is represented by the EEOC, litigation decisions will be significantly more likely to favor the employee than the organization.
Hypothesis 15:	When there is evidence that performance appraisal ratings decreased or were altered after an employee filed a claim or action has been submitted, litigation decisions will be significantly more likely to favor the employee than the organization.
Hypothesis 16:	When the defendant is a public organization, litigation decisions will be significantly more likely to favor the organization than the employee.
Hypothesis 17:	When there is PA evidence documenting satisfactory performance, litigation decisions will be significantly more likely to favor the employee than the organization.

4. METHOD

4.1. Case Selection

Cases for this review were identified by searching Nexis Uni. The following key terms were searched for within cases published in the five-year period from 2014 to 2018: performance appraisal, performance evaluation, performance management, performance review, performance ratings, performance scores, and performance improvement plans. A broad exploratory search yielded more than 10,000 cases, which is the maximum number of cases LexisNexis will extract in one search indicating that there were many cases that fit this search criteria.

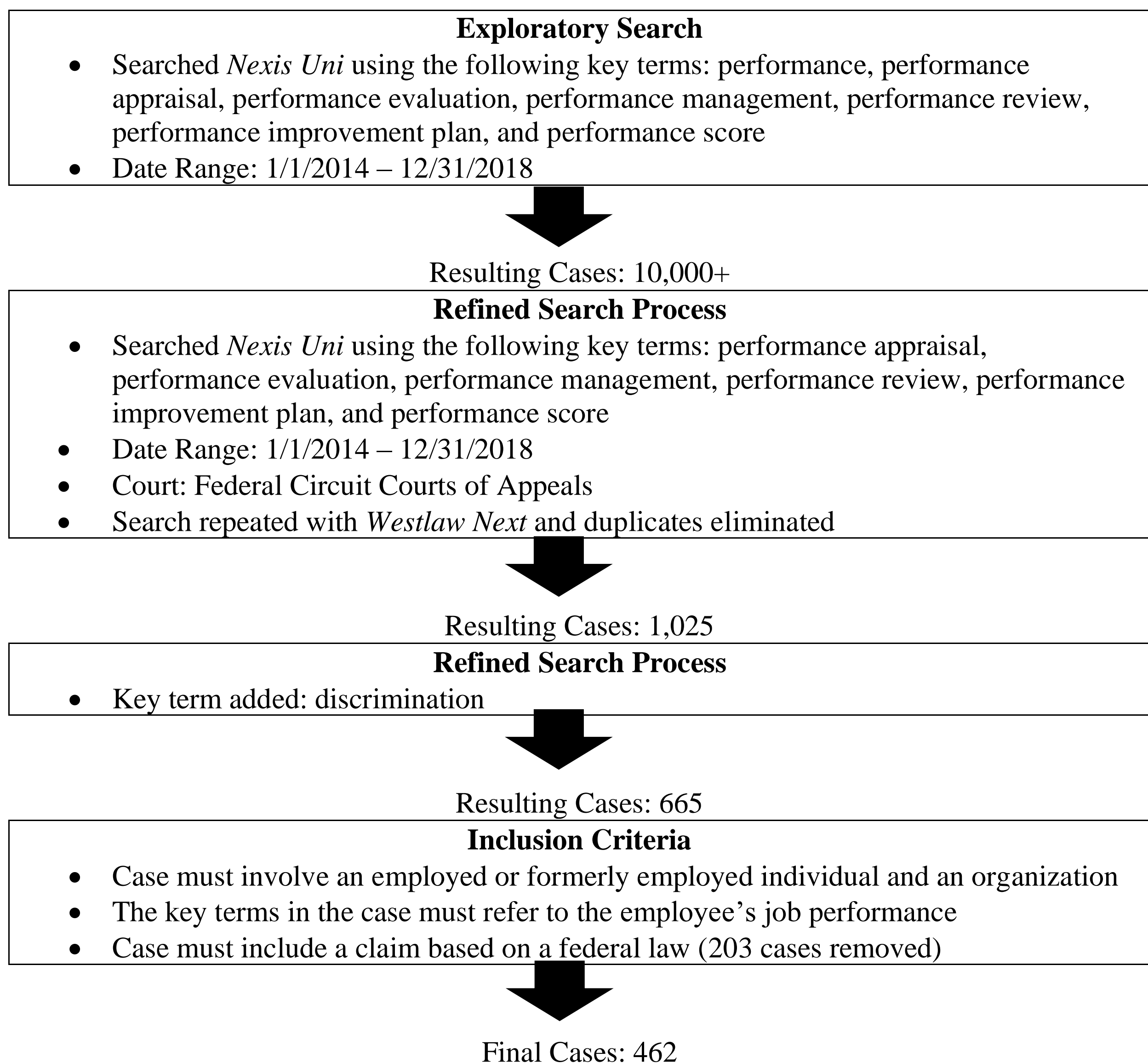
The exploratory search suggested a narrower search was necessary. Correspondingly, the refined search involved the same five years (2014-2018), the same key terms, but was limited to the Federal system at the Circuit Courts of Appeals level, as these cases have higher precedential value relative to lower-level cases. This second search was done using both LexisNexis and Westlaw Next to ensure that all cases fitting these parameters were located. After duplicate cases were eliminated, this search yielded 1,025 cases. In an effort to further reduce the number of cases, the searches were refined with the addition of the key term "discrimination." This addition resulted in 665 cases for consideration.

The author reviewed each case summary, the context surrounding where the search terms appeared in the case test, and screened cases using based on three inclusion criteria. First, the claim must have involved an employee or a formerly employed individual and an organization. Second, any reference to PA or a related term need to have concerned the employee or their coworkers in the focal case. Cases where the key term appeared only in the context of a separate

or precedential case were not included. Third, claims must have been based on federal laws and not state laws. Using these criteria, 203 cases were excluded, resulting in 462 final cases. A depiction of the process and the number of cases identified at each step is illustrated in Figure 2.

Figure 2

Case Identification Flow Chart



4.2. Coding Variables of Interest

Each case was independently read and coded by the author and at least one trained research assistant. Training for research assistants began with a presentation of the study objectives, federal discrimination protections for employees, and the basic structure of case narratives. After this introduction, the author reviewed a detailed coding guide and coding protocol. Cases were coded for variables of interest identified in the proposed hypotheses which were inspired by previous PA case law reviews (e.g., Field & Holley, 1982; Werner & Bolino, 1997). The procedure for coding included documenting the variables in each case using a uniform spreadsheet with variable titles that matched the coding guide. In addition to documenting each variable of interest, coders were required to annotate their copies of the case narratives and highlight the text that supported their coding decisions. All research assistants used these materials and protocol to code six practice cases. The following week the author and research assistants met to discuss the practice cases and compare codes. Discrepancies were discussed until consensus was reached on all cases for all variables of interest.

All cases were coded following a similar iterative process in smaller groups. There were 11 research assistants who were divided into four groups; three groups consisting of three research assistants and one group of two research assistants. The cases were divided between the groups such that each group would code 5-13 cases each week, with the number of assigned cases increasing as coders became more experienced. The author and each member of the group would read, code, and annotate the cases. Once a week, each group met with the author and codes for each variable were compared. Discrepancies between coders were resolved by comparing annotations and textual evidence until complete consensus was reached.

It is important to note that coding was based exclusively on documentation in the case law with the inclination to be more conservative and rely on explicit mention of the variable of interest rather than an inference. It is very likely that the coding is an underrepresentation of the presence of Malos's (1998) content and procedural recommendations in the cases as some of these ideas may not have been explicitly mentioned. For example, a PA may very well have been based on a job analysis, but it would only coded as such if the judge included this information in the case law narrative.

To ensure no variables were overlooked, coders were encouraged to write NA (to mean not available) whenever a variable could not be confidently coded. For example, if there was no mention of PA procedures being communicated formally to the employee, it would not be fair to assume that they were not formally communicated. Instead, this variable was simply coded as NA.

4.2.1. *Discrimination Claims*

Each of the discrimination claims and their respective decisions were noted. There were eight possible Title VII discrimination claims specifying sex, race, national origin, religion, color, disability, age, or veteran status under the CRA. There were also retaliation and harassment claims related to the ADA, PDA and ADEA. Section 1981 discrimination claims were also noted for race, national origin, color, and retaliation. Discrimination and employment-related claims based on state laws were not considered.

Claim Decision. As noted earlier, in the Appellate court, the judges either affirm, reverse, or remand each claim. Because either the organization or the employee can file an appeal, these decisions were coded as either in favor of the employee (1) or in favor of the

organization (0) for each claim. Given the majority of the cases involved more than one claim, a *case* was coded in favor of the employee if at least one claim was found in favor of the employee (e.g., originally found in favor of the plaintiff at trial and affirmed in the Appellate court). This coding is in line with Werner and Bolino's (1997) "decision rule" (p. 12), as well as recommendations by an expert in the field (S. Malos, personal communication, 2018). This was the primary dependent variable for all of the hypotheses; however, additional analyses were conducted for specific claims.

4.2.2. ***Performance Appraisal Content***

Job Analysis. Any mention of the PA being based on a job analysis was coded as present (1). Explicit mention of the PA not being based on a job analysis was coded as not present (0).

Appraisal Content. When available, PA content was coded as: specific (1) or vague (0); controllable (1) or uncontrollable (0); based on behaviors (1) or traits (0); communicated (1) or not communicated (0); and standardized (1) or not standardized (0).

Formal Communication of Appraisal Content. Any explicit text in the case narrative that PA content (e.g., performance dimensions, not ratings) was formally communicated to the employee (1) or not communicated to the employee (0) were coded accordingly.

Standardized and Uniform Appraisal Content. Ideally, PA content is standardized and uniform for all employees within a job group. When this was the case, it was coded accordingly (1) or not (0).

4.2.3. *Performance Appraisal Process*

Formal Communication of Process. Any mention of formally communicating PA procedures to employees was coded as present (1). Alternatively, explicit mention of not formally communicating PA procedures to employees was coded as not present (0).

Multiple, Diverse, Unbiased Raters. The number of unique raters who completed the PAs cited in the case was recorded as a continuous variable. Usually the number of raters who completed ratings on each occasion was consistent but anomalies were recorded.

Written Instructions and Rater Training. The provision of instructions or rater training were both coded as provided (1) or not provided (0).

Performance Documentation and Correction Opportunities. Performance documentation was coded as the total number of PAs presented as evidence in the case. Evidence for correctional opportunities following a PA were coded as provided (1) or not provided (0). Cases were only coded as 'not provided' when the employee was explicitly denied the opportunity to improve following a PA and this was mentioned in the case law narrative (e.g., the employee only received one PA and was fired immediately after). When there was a performance correction opportunity, coding also tracked when the opportunities to improve were formal, through a PIP, or informal.

Review Appraisal Results. Evidence for the opportunity to review the results of the PA was recorded as present (1) or not present (0). Like correction opportunities, reviewing appraisal results was coded as not present when the employee was explicitly denied the opportunity to review their PA and this information was included in the summary.

Formal Appeal and Abuse Detection. As the ability to appeal a PA and the action of appealing a PA are distinct, this recommendation was recorded as two variables. First, the ability to appeal was coded as permitted (1) or not permitted (0) based on information conveyed in the case. Appealing a PA within the organization was coded as appealed (1) or not appealed (0). Cases were coded as a not appealed when the employee declined the opportunity to appeal their PA and this was noted accordingly. Obviously, these variables were related to one another as it would be necessary for there to be the opportunity to appeal in order for an employee to do it.

Frequent Appraisals. The temporal frequency of the evaluations was noted as annual (1), bi-annual (2), quarterly (3), or more frequently than quarterly (e.g., monthly, bi-weekly, and daily; 4).

4.2.4. *Novel Variables*

EEOC Cases. Cases brought forward by the EEOC were dichotomously coded as represented (1) or not represented (0).

Retaliation-Related Events. Three PA-related events associated with retaliation following a protected activity were coded: rating decline, claim of pretext, and claim of fabrication. First, a reduction in any of the PA ratings over time was coded as occurring (1) or not (0). Second, whether the employee claimed they received the lowered or more critical evaluations as pretext was coded as mentioned (1) or not (0). Third, whether the employee claimed the entire PA was fabricated was coded accordingly (1) or not (0).

Public and Private Sector Organizations. The type of organization was coded categorically as private sector (0) or public sector (1).

Satisfactory and Unsatisfactory Evaluations. The presence of satisfactory, unsatisfactory, and mixed evaluations (based on how they were generally described in the narrative) were each coded as present (1) or not (0). The total number of evaluations considered satisfactory and unsatisfactory were also documented. Evaluations with content that reflected both satisfactory and unsatisfactory performance were considered mixed.

Total Recommendations. Malos's (1998) content and procedural recommendations were combined to reflect recommendation adherence with a continuous score. Each recommendation that was adopted was counted as one point. For example, if the PA was based on a JA, it received one point for that recommendation. If more than two raters contributed to a PA, the recommendation for multiple raters was satisfied and received one point for that recommendation. Thus, if all content and procedural recommendations were met, the combined recommendation score would be 14.

4.3. Analysis

For all hypotheses, the dependent variable was the dichotomous case decision indicating the courts either favored the employee or the organization. Thus, the binary logistic regression model is the appropriate statistical test (Stevens, 2009). This form of analysis permits the evaluation of the odds or probability of each litigation decision based on the independent variables. The statistical significance of individual regression coefficients was tested using the Wald chi-square statistic. Goodness-of-fit was assessed with Nagelkerke R^2 due to the disproportionate number of cases with decisions favoring the employee (Nagelkerke, 1991; Peng et al., 2002).

A power analysis (Erdfelder et al., 1996; Faul, et al., 2009) was conducted to estimate the number of cases required to detect the hypothesized effects. Assuming that the odds of a victory for the employee and the organization are equal, a posited small [d = .3] effect, coupled with an alpha of .05 and power of .80 resulted in an estimated 84 cases. Although there is no commonly accepted method for determining a minimum sample size for logistic regression analyses (Aysel, et al., 2019), a minimum of ten cases for each level of the outcome variable (in favor of the employee, in favor of the organization) was adopted (Peng et al., 2002; Tabachnick & Fidell, 2001).

5. RESULTS

Although case decisions are not as simple as winning and losing because they often contain multiple claims, the focal dependent variable in all of the hypotheses was whether the case decision favored the organization or the employee. Correspondingly, each hypothesis was initially tested with the overall decision as the dependent variable. Hypothesis testing within each claim put forward (e.g., discrimination on the basis of sex, harassment, etc.) was also planned, but there was insufficient variability in the outcomes of individual claims for hypotheses testing.

5.1. Claim Information

The appealing party was predominately the employee (445, 96%) and a little over half of the appeals were contesting a decision for summary judgment (no trial necessary) against the employee (237, 51%). The vast majority of cases favored the organization. In total, 380 cases (82%) were found in favor of the organization and 15 (3%) found in favor of the employee. Given there were so few cases favoring the employee, details for these cases are provided in Appendix A. The remaining 67 cases (15%) contained at least one claim that was reversed or remanded back to trial court with instructions to review the employee's claims that were favorable to the employee. Due to the disproportionately fewer cases found in favor of the employee, analyses could not be conducted using the overall appellate decision as the dependent variable. However, analyses could be pursued if the 67 cases that were remanded or reversed were coded as favoring the plaintiff for a total of 82 cases favoring the employee. Arguably, for these cases, there was at least some evidence to support the plaintiff's claim. Further details including the state the cases are appealed from and appeals court are provided in Appendix B.

The average number of claims brought forward by an employee was 3.09 (*SD* = 1.86). The most common claims were Title VII CRA claims (300, 64%), and within those, the most frequent claims were retaliation (230, 77%), followed by claims of race-based (131, 43%) and sex-based (81, 27%) discrimination. The next most common claim was Title VII harassment (71, 24%). Sixty-three percent of the cases (*N* = 190) included both Title VII discrimination and retaliation claims. Claims for the ADEA (85, 18%) and ADA (72, 16%) were less prevalent. There were also a number of cases with claims of Section 1981 (76, 17%) for race-based (62, 82%) and national origin-based (13, 17%) discrimination and retaliation (37, 49%). Further information concerning claims and claim decisions can be seen in Table 5.

5.2. Employment-Related Information

Within the cases examined, the most frequent adverse employment action was termination (334, 72%). The second most frequent claim was an idiosyncratic alleged adverse employment action (82, 17%; e.g., an undesired assignment) followed by a promotion denial (32, 7%), unwanted transfer (7, 2%), and demotion (5, 1%). In roughly half (253, 54%) of the cases, the employee was a woman. On average, employees were employed with the organization for 11.02 years (*SD* = 9.56, Median = 8). The employing organizations were relatively equally distributed across the private- (248, 54%) and public sectors (214, 46%).

With regard to the number of PAs conducted, 215 (46%) cases mentioned only one PA and 247 cases included more than one PA (54%). The number of PAs mentioned in the case's narrative ranged from 1 to 23 (Median = 2; Mode = 1; *M* = 3.23; *SD* = 3.34). The rater's position in the organization was often difficult to determine but in 264 cases it was clear that the supervisor provided at least one of the PA ratings (57%). While specific PA scores were not noted, broader conclusions about the PA content revealed that 311 cases contained unsatisfactory

PAs and 178 (38%) cases contained satisfactory PAs. On average, there were 2.75 (*SD* = 2.72) unsatisfactory PAs and 1.88 (*SD* = 2.21) satisfactory PAs.

Table 5

Frequency and Decisions of Claims Argued in Appeal

Legal Protection	Cases (*N*)	% of Total Cases	% of Specific Claim Type	Decisions in Favor of Org (*N*)	Decisions in Favor of the Employee (*N*)	Outcome Unknown [a] (*N*)
Title VII	300	64		242	6	52
Title VII Retaliation	230	50	77	194	3	33
Sex	81	18	27	68	2	11
Race	131	28	43	117	3	10
National Origin	32	7	11	25	1	5
Religion	10	2	3	9	0	1
Color	4	1	1	4	0	0
Harassment	71	15	24	67	1	3
Family Medical Leave Act	51	11		43	5	4
Americans with Disability Act	72	16		66	2	4
Age Discrimination in Employment Act	85	18		75	1	9
Other EEOC Laws	30	7		25	1	4
Equal Pay Act	7	2	23	6	0	1
Pregnancy Discrimination Act	6	1	20	5	1	0
Veteran	1	.2	3	1	0	0
Rehabilitation Act	16	4	53	13	0	3
Section 1981	76	17		66	4	6
Retaliation	37	8	49	33	1	3
Harassment	10	2	13	10	0	0
Race	62	13	82	52	4	6
National Origin	13	3	17	10	3	0

Note. As plaintiffs could make multiple claims, the number of claims exceeds the number of cases examined. Org = Organization.

[a] Unknown outcomes include claims that were reversed or remanded to a lower court.

5.3. Hypothesis Testing

Many of the hypotheses could not be tested due to a lack of variability in the independent variables. Specifically, Hypothesis 1 (appraisal content based on a job analysis), Hypothesis 2 (specific, controllable behaviors rather than traits), Hypothesis 3 (content communicated to the employee), Hypothesis 4 (standardized and uniform PA), Hypothesis 5 (PA procedures formally communicated), and Hypothesis 7 (written instructions or rater training) could not be tested due to insufficient information in the case narratives. Hypothesis 12 proposed that frequent evaluations would be associated with litigation decisions favoring the organization. While the frequency of the evaluation was available in 168 (36%) cases (138, annual; 30 more frequent than annual), logistic regression could not be conducted as only six of the cases with more frequent PAs were found in favor of the employee. Additionally, only four cases were represented by the EEOC (Hypothesis 14) which provided an insufficient sample to test this hypothesis.

Unfortunately, there was very little information available in the case narratives to indicate if recommendations were followed for three additional hypotheses. Evidence of organizations providing employees with improvement opportunities (Hypothesis 9) was available in 163 (35%) cases. However, there was only one case in which an employee was explicitly denied the opportunity to improve (*Fox v. Leland Volunteer Fire/Rescue Department Inc.,* 2016). There were 280 (60%) cases with evidence that an employee was able to review the results of their appraisals (Hypothesis 10) and yet there were no cases indicating that an employee was denied the opportunity to review their PA. For Hypothesis 11, there was evidence of the opportunity to appeal an appraisal in 87 (19%) cases and evidence of an employee appealing in 72 (16%) cases. There was only one case where an employee had the ability to appeal the appraisal but was

denied the opportunity to do so (*Crawford v. Duke*, 2017). Thus, Hypothesis 9, 10, and 11 could not be tested as proposed. Instead, these hypotheses were tested by comparing cases with evidence that the recommendation was followed to cases in which there was a lack of evidence that the recommendation was followed.

As noted earlier, testing hypotheses within each specific claim was not possible due to insufficient variability in the decisions for sex, race, ADA, ADEA, and Section 1981 claims. However, there was sufficient variability in the 300 (65%) Title VII claim decisions. Of the 300 cases with Title VII claims, 230 (77%) of the cases included an accompanying Title VII retaliation claim. As a case can have both Title VII discrimination claims and Title VII retaliation claims, a chi-square test for association was conducted to assess whether testing hypotheses for these claims separately was appropriate. The relationship between the decisions for Title VII discrimination and retaliation claims was statistically significant ($\chi^2(1) = 180.42$, $p < .05$ and the association was very strong ($\varphi = 0.89$, $p < .05$). Closer examination revealed that most retaliation claims with outcomes in favor of the employee also had Title VII claims with outcomes in favor of the employee. As the outcome variabilities were so closely related, retaliation decisions were not examined independently from Title VII discrimination decisions. Appendix C presents a correlation matrix for the hypothesized independent variables and judicial decisions. Hypotheses were tested first for all of the cases combined (hereafter referred to as the full sample of cases; see Table 6) and then again for just Title VII claims (see Table 7).

Table 6

Logistic Regression Analyses for the Full Sample of Cases

	B	*SE*	Wald statistic	df	Exp (*B*)	*p*	95% CI for Odds Ratio	
							Lower	Upper
Hypothesis 6								
More than One Rater	-0.26	0.31	0.73	1	0.77	.394	0.42	1.41
Number of Raters	0.01	0.06	0.09	1	1.02	.770	0.89	1.12
Hypothesis 8								
Number of Appraisals	0.06	0.05	1.98	1	1.07	.160	0.98	1.16
Hypothesis 9[a]								
Opportunity to Improve	-0.98	0.30	10.85	1	0.38	.001*	0.21	0.67
Hypothesis 10[a]								
Opportunity to Review	0.11	0.25	0.19	1	1.12	.661	0.68	1.82
Hypothesis 11[a]								
(a) Ability to Appeal	-0.14	0.30	0.22	1	0.87	.637	0.48	1.57
(b) Appraisal Appealed	0.02	0.33	.01	1	1.03	.942	1.03	1.97
Hypothesis 13								
Total Recs Followed	-0.03	0.05	.26	1	0.97	.612	0.88	1.08
Hypothesis 16[a]								
Public vs Private Sector	-0.69	0.25	7.85	1	0.50	.005*	0.31	0.81
Hypothesis 15								
Rating Decrease	-0.31	0.28	1.22	1	0.73	.269	0.42	1.27
Hypothesis 17[a]								
Satisfactory Appraisals	-0.93	0.25	14.08	1	2.53	.001*	1.56	4.01

Note. Case outcome was coded as 0 = Organization and 1 = Employee. Recs = Recommendations.
[a] For these categorical independent variables, the comparison referent was the absence of evidence (0).

Table 7

Logistic Regression Analyses for Title VII Claims

	B	*SE*	Wald statistic	df	Exp (*B*)	*p*	95% CI for Odds Ratio	
							Lower	Upper
Hypothesis 6								
More than One Rater	0.85	0.39	4.61	1	2.35	.032*	1.08	5.12
Number of Raters	0.24	0.13	2.89	1	1.25	.089	0.97	1.62
Hypothesis 8								
Number of Appraisals	0.19	0.08	6.44	1	1.22	.011*	1.05	1.42
Hypothesis 9[a]								
Opportunity to Improve	1.49	0.37	16.07	1	4.42	.001*	2.14	9.14
Hypothesis 10[a]								
Opportunity to Review	-0.72	0.3	5.74	1	2.06	.017*	1.14	.3.73
Hypothesis 11[a]								
(a) Ability to Appeal	0.08	0.34	0.05	1	1.08	.816	0.56	2.11
(b) Appraisal Appealed	0.27	0.38	0.51	1	1.31	.474	0.62	2.77
Hypothesis 13								
Total Recs Followed	0.18	0.08	5.7	1	1.19	.017*	1.03	1.38
Hypothesis 16[a]								
Public vs Private Sector	-0.82	0.31	6.84	1	0.44	.009*	0.24	0.82
Hypothesis 15								
Rating Decrease	0.08	0.34	0.05	1	1.08	.814	0.56	2.09
Hypothesis 17[a]								
Satisfactory Appraisals	1.07	0.31	12.01	1	2.92	.001*	1.59	5.34

Note. Case outcome was coded as 0 = organization and 1 = employee. Recs = Recommendations.
[a] For these categorical independent variables, the comparison referent was the absence of evidence (0).

Hypothesis 6 proposed that when employees are reviewed by multiple raters, litigation decisions will be significantly more likely to favor the organization. Information about the number of raters was available in 401 (87%) cases. Within these cases, 295 (74%) organizations relied on one rater and 106 (26%) organizations used more than one rater. When examining the full sample of cases; the presence of multiple raters was not significantly related to an increase in the odds of an organization winning the case, χ^2 (1) = .73, p = .394. Contrary to Hypothesis 6, relying on more than one rater was not more likely to result in outcomes favoring organizations.

When examining just Title VII claims, multiple raters was significantly related to case decisions (χ^2 (1) = 5.01, p = .032), explaining 3.9% of the variance in the overall decision. The change from only one rater (122, 41% of the cases) to more than one rater (82, 27% of the cases) resulted in a 2.35 times increase in the log odds for a decision to favor the organization rather than the employee 95% CI [1.08, 5.12]. Thus, Hypothesis 6 was supported for Title VII cases.

Since the recommendation was for multiple raters, this hypothesis was tested again using a continuous number of raters for the full sample (M = 1.25, SD = 2.60) and Title VII claims (M = 1.27, SD = 2.62). Overall, these additional analyses did not support the use of multiple raters for the full sample (χ^2 (1) = .092, p = .770) or Title VII claims (χ^2 (1) = 3.708, p = .089). Thus, there is limited support for Hypothesis 6.

Consistent and thorough performance documentation was posited to be associated with decisions favoring the organization in Hypothesis 8. Documentation was coded as the total number of PAs presented as evidence. Contrary to expectation, more PAs were not related to the odds of a case decision for the full sample, χ^2(1) = 2.27, p = .160. Although cases favoring organizations tended to have more PAs (M = 3.49, SD = 3.67) than cases favoring employees (M = 2.21, SD = 1.39), the difference was not statistically significant.

The amount of PA evidence in a case was also examined for Title VII claim outcomes. Unlike the results when analyzing all cases, there was a significant relationship between the number of evaluations (*M* = 2.61, *SD* = 1.27) and case decision likelihood (χ_2 (1) = 9.47, *p* = .011) which accounted for 5% of the variance in Title VII claim decisions. For these claims, the number of PAs reported in cases favoring organizations (*M* = 3.49, *SD* = 3.67) was significantly more than the number of PAs reported in cases favoring employees (*M* = 2.20, *SD* = 1.39). Every additional PA within a case was associated with a 1.217 increase in the log odds of a decision favoring the organization over the employee (95% CI [1.04, 1.42]). In other words, for Title VII claims, the more PA evidence, the more likely the organization will win the case.

Hypothesis 9 concerned the provision of an opportunity to improve performance favoring the organization. Further examination of Hypothesis 9 using the presence of a PIP as an opportunity for improvement could not be conducted because only five of the 83 cases with PIP information were decided in favor of the employee.

For all cases, the opportunity for the employee to improve (163, 35%) was significantly related to the outcome of the case (χ_2 (1) = 12.39, *p* = .001) and explained 4.3% of the variance in the decision. Consistent with expectation, an opportunity to improve was associated with a 2.67 times increase in the odds of a decision for the organization (95% CI [1.49, 4.78]). Consistent with the results for the full sample of cases, evidence for the opportunity to improve (126, 42%) was significantly related to Title VII claim outcomes (χ_2 (1) = 19.78, *p* = .001) and explained 10.2% of the variance in these decisions. A documented improvement opportunity was related to a 4.42 times increase in the odds of a decision for the organization (95% CI [0.11, 0.49]). These results provide compelling support for Hypothesis 9. Compared to cases without documentation that the employee was given an opportunity to improve, cases with such

documentation were more likely to be found in favor of the organization.

The employee's ability to review a PA was also hypothesized to increase the odds of a decision in favor of the organization. Hypothesis 10 was examined by comparing the 280 (60%) cases where there was evidence that an employee reviewed their PA to cases where the opportunity to review was not evident (182, 40%). Across all cases, there was no support for Hypothesis 10; the relationship between the ability to review their ratings and the litigation decision was not significant, χ_2 (1) = .19, p = .661.

For Title VII claims, there was a significant relationship between evidence that an employee reviewed the evaluation and the case decision, χ_2(1) = 5.6, p = .001. It explained 3% of the decision variance and evidence that an employee had the opportunity to review their evaluation (210, 70%) resulted in a 2.06 times log odds increase in favor of the organization compared to cases that did not present such evidence (90, 30%) (95% CI [0.11, 0.49]). Whereas Hypothesis 10 was not supported for the full sample of cases, it was supported for Title VII claims.

Hypothesis 11 had two parts. It was predicted that when organizational PA procedures (a) allow employees to appeal PA decisions or (b) there is evidence that an employee pursued an appeal, the litigation decisions will be significantly more likely to favor the organization than the employee. Across all cases, evidence indicating (a) the ability to appeal a PA was found in 87 (19%) cases and (b) PAs were appealed in 72 (16%) cases. All cases where an employee appealed a PA implies the ability to review a PA, meaning 15 (17%) of the cases had employees that did not appeal when they were able to do so.

Hypotheses 11a and 11b were tested separately for the full sample of cases and Title VII claims. Neither the ability to appeal (χ_2 (1) = .22, p = .637) nor the action of appealing (χ_2 (1) =

.01, p = 942) was significantly related to the decisions for all cases. Similar to these findings, evidence of the ability to appeal a PA (76, 25%; χ^2 (1) = 0.05, p = .816) and evidence that an employee appealed their PA (62, 21%; χ^2 (1) = 0.53, p = .474) were not significantly associated with the Title VII claim decisions. Therefore, there was no support for Hypotheses 11a or 11b.

Given the wide range of Malos's PA content and procedural recommendations, it was expected that organizations that follow more of these recommendations will fare better than organizations that follow less of these recommendations (Hypothesis 13). A continuous measure of adherence to recommendations (M = 1.19, SD = 0.61; Min = 0, Max = 9) was not significantly related to the odds of the case decision for all cases examined, χ^2 (1) = .26, p = .612. Although on average, cases favoring the organization (M = 1.22, SD = 0.61) followed more recommendations than cases favoring employees (M = 1.08, SD = 0.60), the difference was not statistically significant.

Whereas Hypothesis 13 was not supported for the full sample, it was supported for Title VII claims. There was a significant relationship between the total number of Malos's recommendations followed in a case (M = 1.22, SD = 0.63) and the odds of the decision (χ^2 (1) = 6.04, p = .017), explaining 3.2% of the variance in these decisions. The total number of recommendations followed was significantly higher for cases found in favor of organizations (M = 1.26, SD = 0.62) than cases found in favor of employees (M = 1.04, SD = 0.65). For every additional recommendation followed, there was a 1.19 times increase in the odds of a decision favoring the organization (95% CI [1.03, 1.38]). In other words, the more of Malos's recommendations that an organization follows, the more likely the courts will find in their favor.

It was also predicted that public sector organizations would fare better than private sector organizations because they would generate more documentation to support their decisions

(Hypothesis 16). There were 248 (53.7%) cases with claims against a private sector organization and 214 (46.3%) with claims against a public sector organization. Organization type was significantly related to the case decision (χ^2 (1) = 8.02, p = .005) and explained 2.8% of the decision variance. Contrary to expectation, the log odds of a case having a decision for the organization is .5 times lower for public sector organizations compared to private sector organizations (95% CI [0.31, 0.81]). These results do not support the hypothesis that public sector organizations fare better than private organizations in litigation involving PA. In fact, the results indicated that private sector organizations tend to fare better.

The same results were found for Title VII claims. Once again, there was a significant relationship between the organization type and case decision (χ^2 (1) = 7.24, p = .009; Nagelkerke R^2= .038). Public organizations (160, 53%) had a 0.44 decrease in the odds of a favorable decision compared to private organizations (140, 47%; 95% CI [0.24, 0.82]). Consistent with the results for the full sample of cases, the results for Title VII claims were the opposite of what was predicted. Contrary to Hypothesis 16, private sector organizations were more likely to win than public sector organizations.

5.3.1. ***Hypothesized Decisions for Employees***

Many case narratives reported changes in an employee's PA after they engaged in a protected activity. While only seven cases claimed that a PA was altered or fabricated (e.g., Burton v. Freescale Semiconductor; see Appendix A), in 97 (21%) cases there was evidence of a decrease in PA ratings following an employee's participation in a protected activity. Hypothesis 15 proposed that when there is evidence of this combination of actions, litigation decisions are more likely to favor the employee than the organization. However, a drop in PA ratings after a protected activity was not significantly related to the litigation decisions for the full sample of

cases (χ_2 (1) = 1.19, p = .269). Likewise, there was not a significant relationship between a decrease in rating (74, 25%) and Title VII decisions (χ_2 (1) = 0.05, p = .814). Thus, no support was found for Hypothesis 15; evidence of a decrease in PA ratings following an employee's engagement in a protected activity was not related to case decisions in favor of the employee.

Hypothesis 17 proposed that when there is evidence of satisfactory performance in PAs, litigation decisions will be more likely to favor the employee. The presence of satisfactory PA evidence (178, 38%) was significantly related to the case decisions (χ_2 (1) = 14.31, p = .001), explaining 0.5% of the variance in the overall decision. When there is evidence of satisfactory performance, the odds of a case resulting in a decision favoring the employee (48, 27%) is 2.53 times higher than the odds of a case resulting in a decision favoring the organization (130, 73%), 95% CI [1.56, 4.01].

There was also a significant relationship between evidence of satisfactory performance ratings (136, 45%) and Title VII claim decisions (χ_2 (1) = 12.75, p = .001). The relationship explained 6.7% of the decision variance. There was a 2.92 times increase in the log odds of a decision for the employee when comparing cases with no evidence of satisfactory PA, (161, 54%) to cases where there is evidence of a satisfactory PA 95% CI [1.59, 5.34]. Thus, these results support the hypothesis that employees are more likely to win a claim when there is evidence of satisfactory performance.

Overall, more hypotheses were supported when examining just Title VII claims compared to the full sample of cases. Only two hypotheses were supported for the full sample. Opportunity for the employee to improve performance (Hypothesis 9) was associated with more decisions favoring the organization, and evidence of satisfactory performance (Hypothesis 17) was associated with decisions favoring the employee. Contrary to Hypothesis 15, the courts were

more likely to rule in favor of private sector organizations than public sector organizations regardless of the claim. When examining just Title VII claims, multiple raters (Hypothesis 6), more rather than less PA documentation (Hypothesis 8), and the opportunity for the employee to review appraisal results (Hypothesis 10) were related to decisions favoring the organization. A summary of the results of hypothesis testing for both samples of cases is provided in Table 8.

Table 8

Summarized Results of Tested Hypotheses

	Hypothesis (H)	Statistical Support for all claims examined	Statistical Support for Title VII claims
H6:	When employees are reviewed by multiple raters, litigation decisions will be significantly more likely to favor the organization than the employee.	Not Supported	Supported
H8:	When organizational PA procedures require thorough and consistent performance documentation, the litigation decisions will be significantly more likely to favor the organization than the employee.	Not Supported	Supported
H9:	When organizational PA procedures note deficiencies and provide employees with an opportunity to correct these deficiencies, litigation decisions will be significantly more likely to favor the organization than the employee.	Supported	Supported
H10:	When employees are allowed to review their ratings, litigation decisions will be significantly more likely to favor the organization than the employee.	Not Supported	Supported
H11:	When organizational PA procedures (a) allow employees to appeal PA decisions, and/or (b) there is evidence that an employee pursued an appeal, the litigation decisions will be significantly more likely to favor the organization than the employee.	Not Supported	Not Supported

Table 8 Continued

	Hypothesis (H)	Statistical Support for all claims examined	Statistical Support for Title VII claims
H13:	When there is evidence that an organization followed a greater number of Malos's content and procedural recommendations, the litigation decisions will be significantly more likely to favor the organization than the employee.	Not Supported	Supported
H15:	When there is evidence that performance appraisal ratings decreased or were altered after an employee filed a claim or action has been submitted, litigation decisions will be significantly more likely to favor the employee than the organization.	Not Supported	Not Supported
H16:	When the defendant is a public organization, litigation decisions will be significantly more likely to favor the organization than the employee.	Not Supported[a]	Not Supported[a]
H17:	When there is PA evidence documenting satisfactory performance, litigation decisions will be significantly more likely to favor the employee than the organization.	Supported	Supported

Note. A $p < .05$ cutoff was used when determining support for each hypothesis.
[a]The relationship was statistically significant but the direction of the log odds change was contrary to the hypothesis.

5.4. Exploratory Analysis

In addition to the previously examined recommendations and hypothesized elements, other variables were also extracted from case narratives. Exploratory analyses focused on more nuanced assessment of employee performance reflected in the PA: the total number of satisfactory evaluations and the presence of unsatisfactory evaluations, as well as the total number of unsatisfactory evaluations. Additionally, employee tenure (years on the job) was also examined as employees with longer tenure are likely to have more PA evidence.

The valence of evaluations were coded based on how they were described in the case. Only those that were clearly satisfactory or unsatisfactory were coded accordingly. Evaluations that contained evidence of both were coded as mixed. For example, the narrative of *Wheat v. Fla. Parish Juvenile Justice Comm'n*, 2016, included a reported PA rating of "Very Good" and was considered to be a satisfactory PA. The case narrative of *Frakes v. Peoria Sch. Dist. No. 150* (2017) included evidence of one unsatisfactory evaluation. Evidence of unsatisfactory performance for the former special education teacher's PA included the overall "unsatisfactory" rating and documented performance deficiencies noting issues with classroom management, tardiness, and a dearth of faculty presentations. The narrative of *Burton v. Freescale Semiconductor, Inc*. (2015) contained evidence of a mixed PA because the judge described the 2009 and 2010 PAs as "neutral" containing positive and negative ratings of performance.

The likelihood of case outcomes and the continuous number of positive evaluations from the 178 cases with satisfactory PA evidence were examined first. It was expected that more satisfactory PA evidence would result in more decisions for employees due to the relationship between cases with evidence of at least one satisfactory evaluation and decisions favoring the employee. However, there was not a significant relationship between the number of satisfactory PAs presented (M = 1.88, SD = 2.21) and case decisions for the full sample of claims ($\chi^2(1)$ = .076, p = .780). This indicates that the number of positive PAs was unrelated to the odds of decisions in favor of the organization or the employee. Likewise, there was not a significant relationship between the number of satisfactory PAs and Title VII case decisions (M = 1.85, SD = 2.04; $\chi^2(1)$ = 1.54 , p = .311). Whereas the presence of satisfactory evaluations was associated with a higher likelihood of decisions for the employee, the quantity of satisfactory evaluations was not.

Exploratory analyses were also conducted on the presence of unsatisfactory evaluations. Given the positive relationship between satisfactory evidence and decisions for employees the opposite relationship was expected for unsatisfactory PA evidence. Unsatisfactory evaluations appeared in a majority (311, 67%) of the cases examined and among cases with Title VII claims (227, 76%). There were no significant relationships between case decisions across the full sample of cases ($\chi^2(1) = 1.68$, $p = .197$) nor cases with Title VII claims ($\chi^2(1) = 1.85$, $p =.165$). Comparing cases with evidence of unsatisfactory performance to those with no evidence of unsatisfactory performance did not result in any significant differences in the odds of case decisions.

The amount of unsatisfactory PA evidence was also compared to case decisions. Unlike the results for the number of satisfactory evaluations, there was a significant relationship between the number of unsatisfactory evaluations and the odds for case decisions when examining the 311 cases with unsatisfactory PA evidence ($M = 2.75$, $SD = 2.72$; $\chi^2(1) = 14.96$, $p = .003$). Title VII claim decisions were also significantly related to the quantity of negative evaluations ($M = 2.75$, $SD = 2.49$; $\chi^2(1) = 20.18$, $p = .002$). These relationships explained a greater amount of the decision variance (7.5% and 14%, respectively) than the presence of satisfactory evaluations. Each additional unsatisfactory PA resulted in a .72 and .54 times decrease in the log odds of a decision for the employee (95% CI [.58, .89]; 95% CI [.37, .79]). Whereas decisions for the employee were not related to the presence (or absence of) unsatisfactory PA evidence, decisions were less likely to be found in the employee's favor when there was more rather than less unsatisfactory PA evidence.

While it may appear odd that the dichotomous and continuous variables measuring satisfactory and unsatisfactory evidence have contradictory results, the case decisions being

compared are different. The dichotomous variables compare the presence of evidence or the lack of evidence among all 462 (or 300 for Title VII) cases. Analysis of the continuous variables compare the likelihood of case decisions among a smaller number of cases that contained any evidence.

Another variable examined in an exploratory fashion was employee tenure (351, 76%). Contrary to speculation, employee tenure was not significantly related to the number of PAs presented as evidence (r (350) = .08, p = .156). There was not a significant relationship between tenure and case decision for the full sample of cases ($\chi^2(1) = .152$, $p = .695$) or Title VII claim decisions ($\chi^2(1) = .104$, $p = .745$). The total amount of time an employee worked for an organization was not significantly related to the amount of PA evidence or the case decision.

Table 9 presents all of the exploratory analyses conducted. Exploratory analyses of the continuous number of satisfactory PAs revealed that, while the presence of PA evidence may have been related to higher odds of favorable employee decisions, the amount of PA evidence was not significantly related to these odds. The presence and quantity of unsatisfactory PA evidence had the opposite finding. The presence of unsatisfactory PA evidence was unrelated to case decisions alone but the quantity of unsatisfactory evidence was related to decisions in favor of the organization. The likelihood of decisions for the organization increased when there was more evidence of unsatisfactory PAs. Finally, employee tenure did not have a significant influence on case outcomes.

Table 9

Exploratory Logistic Regression Analyses

Variable	*B*	*SE*	Wald statistic	df	Exp (*B*)	*p*	95% CI for Odds Ratio	
							Lower	Upper
All Case Decisions								
Number of Sat Eval	-.02	.07	0.08	1	1.02	.780	0.88	1.18
Presence of Unsat Appraisals [a]	.36	.27	1.78	1	1.44	.197	0.84	2.45
Number of Unsat Appraisals	-.33	.11	1.08	1	-0.72	.003*	0.58	0.89
Tenure	.01	.02	0.15	1	1.01	.695	0.98	1.04
Title VII Case Decisions								
Number of Sat Appraisals	-.15	.14	1.03	1	0.87	.311	0.65	1.15
Presence of Unsat Appraisals [a]	-.43	.32	1.74	1	0.65	.165	0.35	1.23
Number of Unsat Appraisals	-.61	.19	9.87	1	-0.51	.002*	0.37	0.79
Tenure	.01	.02	0.11	1	1.01	.745	0.97	1.04

Note. Case decisions were coded as 0 = organization and 1 = employee.
Eval = Evaluations, Sat = Satisfactory, Unsat = Unsatisfactory
[a] For these variables, the comparison referent was the absence of evidence (0).
* $p < .05$

6. DISCUSSION

PAs are a pervasive and arguably notorious form of performance management. While many have questioned the value of PA, the measurement and evaluation of employee performance is a critical component of HR management. PAs serve many developmental and administrative purposes within organizations. Administrative purposes include high-stakes decisions about an employee's future at an organization. Because PAs are often the only formal documentation an organization has of an employee's performance, it is reasonable to assume PA content and process information would appear in case law summaries that reference this evidence.

Prior to quantitative examinations, Cascio and Bernadin (1981) reviewed the extant PA literature to determine what PA content and process characteristics were most commonly recommended and supported. These characteristics and recommendations have expanded with each subsequent examination and have culminated in the six PA criteria and nine PA process recommendations consolidated by Malos (1998; 2005).

Whereas the legal protections against discrimination have not changed, the interpretation of these laws and the way PAs appear in litigation may have. The purpose of this study was to provide an updated review of the role of PA in employment discrimination litigation by systematically examining the extent to which Malos's recommendations appear in 462 U.S. appellate employment discrimination cases that mention performance appraisal and took place over the past five years (2014-2018). Previously examined PA characteristics were included in the coding guide alongside characteristics and features of the PA that were inspired by contemporary PA literature (e.g., Gorman et al., 2017). In an effort to extend previous

examinations, alternative operationalizations of PA characteristics were pursued whenever possible including continuous, rather than dichotomous, measures. In the process of reading and coding these cases, new insights about the presence of PA in litigation also emerged.

6.1. Malos's Recommendations

Twelve formal hypotheses were derived from Malos's (1998) recommendations. Unfortunately, due to a paucity of information within the case narratives, six hypotheses (job analysis, measuring specific controllable behaviors , communicating appraisal content, standardized PAs, communicating appraisal procedure, and providing specific instructions or training for raters) could not be tested. A seventh hypothesis (appraisal frequency) could not be tested due to the lack of variability in the case decisions. Ultimately, five hypotheses based on the recommendations and four novel hypotheses were tested twice – once with the full set of 462 cases and once with a subset of 300 Title VII cases.

Within the full sample of cases, only two hypotheses were supported. Only one of these hypotheses was based on a Malos (1998) recommendation. Consistent with Barrett and Kernan's (1987) findings, evidence of the opportunity to improve performance was associated with more decisions favoring the organization. Providing the opportunity to correct deficient performance was proposed to contribute to perceptions of fairness. Organizations that permit developmental opportunities could be perceived as more procedurally just.

More hypotheses based on Malos's (1998) recommendations were supported when limiting the cases to just the Title VII claims. Within this subset of claims, the provision of opportunities to improve performance, multiple raters, consistent and thorough PA documentation, and the opportunity for the employee to review appraisal results were related to decisions favoring the organization. These findings mirror three of Malos's procedural

recommendations that have been previously examined in quantitative analyses (e.g., Werner & Bolino, 1997). These procedures were expected to provide a stronger corroborative defense of the organization's administrative decisions as more raters and more documentation would provide more data on which to make decisions. Finally, it is unreasonable to expect an employee to improve their performance if they are unaware of their performance deficiencies. As an extension, keeping this information from the employee also denies them the opportunity to improve their performance. Given the strong association found between opportunities to improve and other procedure-related recommendations, it appears that these practices remain important for a legally defensible PA process. Table 10 summarizes the recommendations that could not be tested and the results of those that could.

6.2. Comparisons to Previous Examinations

The inconsistencies between the results of this examination (see Table 11) and previous examinations are noteworthy given the many methodological similarities. Perhaps the previous review most similar to the current study is Werner and Bolino's (1997) examination of 295 appellate cases between 1980 and 1995. In both studies, cases were gathered using a keyword search that included "discrimination" of employment litigation in legal databases. The search and selection of discrimination cases was done strictly at the appellate level from the U.S. Courts of Appeals circuit using the approach taken by Werner and Bolino (1997). Further, cases were only included in the final sample if the PA was central to the case itself. Case decisions that were coded as in favor of the organization or the employee and analyzed using logistic regression. Also, the coding of the decision outcome dependent variable was informed by Werner and Bolino's (1997) "decision rule" (p. 12), as well as personal correspondence with Malos (S.

Malos, personal communication, 2018). Both studies found that multiple raters and opportunity for the employee to review the appraisal were related to decisions favoring the organization.

Table 10

Summary of the Examined Performance Appraisal (PA) Recommendations

Malos's (1998; 2005) Recommendations
PA Criteria should be…
1. objective rather than subjective
2. job-related or based on a job analysis
3. focused on behaviors rather than traits
4. within the control of the ratee
5. specific functions, not global assessments
6. communicated to the employee
The PA process should…
1. be standardized and uniform for all employees within a job group
2. be formally communicated to employees
3. use multiple, diverse, and unbiased raters
4. provide written instructions and training to raters
5. contain thorough and consistent documentation across raters that include specific examples of performance based on personal knowledge
6. provide employees notice of performance deficiencies and opportunities to correct them
7. provide access for employees to review appraisal results
8. provide formal appeal mechanisms that allow for employee input
9. establish a system to detect potentially discriminatory effects or abuses of the system overall

Note. Recommendations in grey could not be tested. Recommendations in green were supported when examining all cases and Title VII claims. Recommendations in orange were supported only when examining Title VII discrimination claims. Recommendations in red were not supported by either examination.

Table 11

Comparisons to Previous Quantitative Examinations

Recommendation	Citation for Study				
	Feild & Holley (1982)	Feild & Thompson (1984)	Miller et al. (1990)	Werner & Bolino (1997)	Mendoza (2020)
The PA process should…					
1. Be standardized & uniform					Insufficient Evidence
2. Formally communicated					Insufficient Evidence
3. Multiple, diverse, & unbiased raters	Not Supported			Supported	Partially Supported
4. Instructions & training	Supported	Supported	Insufficient Evidence	Supported	Insufficient Evidence
5. Documentation	Not Supported				Not Supported
6. Opportunity to improve					Supported
7. Review results	Supported	Supported	Insufficient Evidence	Supported	Partially Supported
8. Allow for Appeal					Not Supported
9. System to appeal PA abuse					Not Supported

Note. Descriptions of performance appraisal (PA) characteristics were standardized across studies. Empty cells indicate a recommendation was not considered, supported or not supported indicates whether the recommendation was significantly related to decisions favoring organizations, insufficient evidence reflects when Information about the recommendation was recorded but there was insufficient evidence for analysis.

Whereas there was considerable similarity in the methods used in the current study and the methods used in Werner and Bolino's (1997) study, there were several notable differences in the results. In this previous examination, case decisions were more evenly distributed in favor of the organization (58%) and the employee (42%). In contrast, the case decisions for this sample were overwhelmingly in favor of the organizations (82%) with very few decisions entirely in

favor of the employee (3%) and a large number of reversed or remanded decisions (15%). In fact, the distribution of decisions was so uneven that there was insufficient variability to test some of the hypotheses. Werner and Bolino (1997) found that basing the PA on a job analysis, providing written instructions to raters, rater agreement, and allowing employees to review the results of their PAs were significantly related to decisions favoring the organization (Werner & Bolino, 1997). Insufficient information was available to calculate agreement across multiple raters, the extent to which the PA was based on a job analysis, and the provision of written instructions/training could not be examined. The inability to test these recommendations and the disparity in the ratios of case decisions was surprising given the methodological improvements that were made to address the limitations identified in previous studies which will be elaborated upon next.

6.2.1. *Number of Cases Examined*

A number of methodological enhancements were made to this study in an effort to address a number of limitations identified in former reviews. The first enhancement was the greater number of cases examined). Almost all of the previous quantitative examinations consisted of less than 70 cases (Feild & Holley, 1982; Feild & Thompson, 1984; Miller et al., 1990). Each of these previous reviews identified sample size as a limitation and suggested that a larger number of cases would be necessary to support the existence of the relationships between PA recommendations and case decisions. Prior to the present review, the largest quantitative examination of case law was Werner and Bolino's (1997) review of 295 cases. Thus, the current review consists of the largest number of cases examined in one study.

Previously, the following three PA process recommendations were supported: reviewing appraisal results, notice of performance deficiencies and correction opportunities, and multiple

raters. Only the notice of performance deficiencies and correction opportunities recommendation was supported across all cases in the current study. In contrast, all three recommendations were supported when examining only the smaller subset of Title VII cases.

6.2.2. ***Missing Data***

A very interesting difference between this examination and its predecessors is the number of PA characteristics that could not be located in the narratives of the cases examined. Several of the previous reviews cite missing data as the largest issue facing quantitative analyses of litigation (Feild & Holley, 1982; Werner & Bolino, 1997). There was insufficient PA information to test any of the content-based recommendations put forth by Malos (1998, 2005). Previously, quantitative reviews found PA *criteria* that were based on a job analysis, focused on behaviors rather than traits, and were within the control of the ratee were significantly related to decisions for organizations (Feild & Holley, 1982; Feild & Thompson, 1984). Information to test two previously supported PA *process* recommendations (Feild & Holley, 1982; Feild & Thompson, 1984; Werner & Bolino, 1997) was also not available for the provision of written instructions or using standardized/uniform procedures. Given the number of cases reviewed in this study, the absence of evidence for these PA recommendations is surprising and raises questions about the role these recommendations play in more recent litigation.

Another possible reason for the discrepancy in results may be the violation of heteroscedasticity when analyzing data with very small sample sizes. Given previous reviews consisted of a smaller number of cases, examination of these rare events of explicit deviations from PA recommendations often involve a violation of this assumption. For example, Werner and Bolino's (1997) finding that organizations received more favorable decisions when employees had the ability to review their PA was based on the result of chi-square analyses with

only three decisions in the cell where an employee was not allowed to review their PAs and the decision was in favor of the organization. Indeed, comparisons regarding the use of a job analysis to identify appropriate PA criteria were made when there were only five cases where decisions favored the employee and a job analysis was not utilized (Werner & Bolino, 1997).

6.2.3. *Possible Explanations of Differences*

Court Level and Case Decisions. First, the level of the court could have a significant effect on the decisions that are made in employment litigation or discrimination cases. With the exception of Werner and Bolino (1997), other examinations considered trial and appellate decisions together. Although appellate cases represent only a subset of litigation and are likely to have less details than trial cases, the decision to focus on appellate cases like Werner and Bolino was made because they are more likely to serve as precedent in future litigation.

The proportion of case decisions in favor of organizations could be due to the focus on appellate cases and the party who initiated the appeal. In the current study, the appealing party was almost always the employee (96%). While the appealing party for the previous quantitative examinations is not noted, there is some literature that suggests appellate judges are unlikely to overturn a lower court's decision. Indeed, while the rate varies by the type of case, a majority of all appealed decisions are reversed (Carp et al., 2019; Guthrie & George, 2004). Thus, the proportion of appeals brought forward by employees in this sample and the rate at which appeals are overturned could contribute to the differences between this examination and its predecessors.

There are two proposed explanations for the small number of cases overturned on appeal. First, the trial decisions are unlikely to be overturned because trial court judges are shaping their decisions on the precedent set by higher courts (De Mesquita & Stephenson, 2002). Given the decades that have passed since the CRA was passed, it is quite likely that there is a precedent

that could apply to these cases. Indeed, one example of precedent related directly to PAs was established in the case of *Lucas v. WW Grainer, Inc*. (2001) where it was decided that, on their own, a negative performance evaluation does not constitute an adverse employment action. Trial judges can refer to this precedent in any future cases where an employee alleges that the alleged adverse employment action was a PA.

The second explanation for smaller number of overturned decisions in favor of the employee could also be a biased predisposition of appellate judges to agree with trial decisions. A judge's subconscious decision-making processes including confirmation bias (see Oswald & Grosjean, 2004 for a review) could interfere with their perception of trial case decisions. A literature review of the factors that influence appellate decisions revealed that the previous decision explains the appellate decision more than any other considered factor (Boyd et al., 2010). In an experimental study, Edwards (2018) measured the extent to which knowledge of a prior decision could affect law students' legal judgments when all other factors are held constant. The law students were more likely to agree with prior decisions when they were known. Edwards proposed that if judges behaved similarly the affirmation rate in appellate courts could be inflated by as much as 8% due to a bias in favor of affirming prior decisions.

While these explanations help to explain the high rate of trial decision confirmation, they do not explain the differences in the hypotheses supported. Werner and Bolino's (1997) examination focused only on appellate cases but they did not report the proportion of appeals made by employees or organizations. It could be that a greater number of appeals in the current study came from employees. The change in proportion of the appealing party could even reflect larger changes in employment litigation and the defense of discrimination claims. However, without knowing the ratio of the appealing party, no comparisons of the composition of

appellants can be made. Comparisons to Feild and Holley (1982) and Feild and Thompson (1984) are even more difficult due to the mix of trial and appellate cases.

Comparing Claims. This review examined a much larger number and types of claims than any of the previous reviews. Some of the considered legal discrimination protections did not yet exist at the time of previous reviews. As noted earlier, more hypotheses were supported for the subset of Title VII cases. Comparisons of how PAs and their recommended content and process characteristics appear across claims of ADA, ADEA, PDA, and Title VII could be comparing apples to oranges.

It may be likely that the claims are too different from each other to be able to meaningfully extract how PAs and PA characteristics appear across all federal discrimination protections. Each legal protection can differ in a number of ways that could have influenced the likelihood of case decisions in the complete sample. In a quantitative examination of only ADEA discrimination claims, Miller, Caspin, and Schusater (1990) noted that PA characteristics found in previous examinations of discrimination claims were not reported in their sample of cases.

It may also be that the similarities in the Title VII cases is what allowed the relationships between the recommendations and the case outcomes to emerge. While sex and race claims of discrimination are different, they require the same burdens of proof and can be compared more directly. Still, even within Title VII protections, employees are provided different protections for claims of discrimination and retaliation. Disparate treatment discrimination against an individual is prohibited for any aspect of employment while retaliation can only occur after engaging in a protected activity and employees must show that they experienced an adverse action.

Availability of Appraisal Information. The absence of the job analysis recommendation for criteria is particularly interesting because it was found to be significantly

related to case decisions for organizations across almost all quantitative reviews (Feild & Holley, 1982; Feild & Thompson, 1984; Werner & Bolino, 1997). While evidence of several process recommendations was also unavailable, there was considerably more process-related evidence compared to content-related evidence.

The relative dearth of extractable evidence for these PA criteria recommendations compared to process recommendations could be a function of the level of involvement an employee has in the process of the PA's development or administration. Employees are much more likely to experience process-related recommendations compared to content recommendations. While it is very unlikely that an employee would be involved in development and selection of the criteria evaluated in their PAs, employees could directly observe the adherence (or non-adherence) to PA process recommendations when receiving their own evaluations of performance. These observations may be relevant to the employees claim and thus be more likely to be mentioned in the case narratives.

It is also important to note that many of the discrimination claims do not directly involve the PA itself. Given that providing and reviewing PA information is not the objective of the court or the case narratives, the PA information that is more likely to be included (or excluded) in case narratives can be telling. Information about how the PA criteria was selected and whether the criteria was appropriately developed might only occur in cases where the PA itself is under review. Further, because negative PAs themselves are not considered adverse employment actions (*Lucas v. WW Grainer, Inc.,* 2001), the details of the PA criteria might receive less attention even when an employee is alleging that the content was an unfair measure of their performance. Because case narratives are written by judges and these judges are responsible for summarizing the facts of the case and the rationale behind the case's decision, judges may be

more likely to focus on other alleged employment actions that could qualify and support claims of discrimination.

Information about the organizational procedures that are followed could also be pertinent to the support or rebuttal of discrimination claims. Most cases did not provide much detail about the specific scores or ratings an employee received on their PAs. Instead the case narratives were more likely to reveal a general summary of whether the performance reflected in the PA was satisfactory or unsatisfactory. This evidence was used by both employees and organizations making it important evidence for a judge to mention in the case narrative.

6.3. Contributions

In addition to Malos's recommendations, four new hypotheses were tested. First, as an extension of the practices previously identified important for a legally defensible PA process, the effect of the number of followed recommendations was examined. This examination revealed that following more of Malos's recommendations was related to a greater likelihood of decisions for organizations across the full set of cases and the Title VII subset.

Second, and contrary to what was expected, the court's decisions were more likely to be in favor of private sector organizations than public sector organizations regardless of the claim. Private and public sector organizations are beholden to different expectations of justification and notice when terminating employees. Due to the need to follow more regulations, public sector organizations are held to a higher level of accountability and they must provide justification to support their administrative decisions. Establishing and supporting a cause for termination necessitates the collection of performance-related or PA documentation. In addition to more PA documentation, the establishment of cause was also believed to be related to PA process recommendations for correction opportunities, access to PA results, and formal appeal

mechanisms. It was proposed that the higher requirements of justification would make public sector organizations more likely to adhere to these PA process recommendations and, in turn, be more likely to successfully defend their administrative decisions. However, it appears that the opposite might be true. While documentation of employee performance is also done in private sector organizations, the need is not necessarily as great as it is for public sector organizations. Feasibly, whatever associated benefits the PA process-related recommendations may have provided could be mitigated by the higher burden of justifications required by public sector organizations.

The nature of the performance captured in the PAs presented as evidence and the odds of decisions in favor of the employee rather than the organization was the focus of the last two exploratory hypotheses. There was no support that evidence of decreasing performance appraisal ratings following a protected action was related to decisions in favor of the employee. This was proposed to be representative of a potential way in which retaliation against an employee could be taken.

There was evidence to support the hypothesis that the presence of satisfactory performance was associated with outcomes favoring the employee in both samples. Whereas an employee's performance can certainly change over time, it was believed that evidence of satisfactory performance would be important for employees who have claimed a discriminatory motive behind an alleged adverse employment action. If there was no evidence of satisfactory performance presented in the case narrative it could be much harder for an employee to argue that the adverse employment action was not performance-based.

Interesting relationships emerged when comparing of the presence of satisfactory and unsatisfactory PA evidence and the relative amount of each type of evidence. While the presence

of satisfactory evaluations was related to an increase in the employee's odds of a favorable decision, the quantity of satisfactory evaluations was unrelated. In contrast, evidence of unsatisfactory PAs was unrelated to the odds of case decisions but decisions were less likely to be in the employee's favor when there was more rather than less unsatisfactory PA evidence. It would appear that the amount of satisfactory PA information is not as important as providing some evidence that performance was satisfactory at one point. In contrast, the presence of unsatisfactory performance is less important than the amount of unsatisfactory performance.

6.3.1. ***The Employee's Perspective***

It is important to contextualize that the cases examined represented situations where an employee feels they were wronged so egregiously that they pursue litigation. In employment litigation, there are substantial hurdles that employees making claims must overcome for a case to result in a decision or judgment. In fact, cases that go to the trial court are considered anomalies (Estreicher et al., 2018). Cases that go to appeals are even rarer. Still, examining the PA information available in these appeal cases can reveal important trends.

Ultimately, the employees in the examined cases are pursuing their claims of discrimination despite a great financial investment, the long process of litigation, and a huge difference in resources compared to organizations. Previous reviews of legally defensible PA characteristics in discrimination cases have focused largely on the perspective of the employer and have not addressed the hurdles employees face in the process of defending their claims. Given that almost all of these appeals were made by employees and almost half of these were contesting summary judgments against their claims, the fact that any employees prevailed at all is compelling and worth examination.

Navigating the process from receiving the EEOC "right to sue" to a courtroom requires knowledge of the legal process that employees may not have. Legal counsel in employment discrimination often utilizes a contingent fee structure where the employee does not have to assume most of the upfront costs of raising a claim in exchange for an agreement that attorney's fees will be paid out in the event of a case victory (Donohue & Siegelman, 1991). While employees often seek out these agreements with experienced counsel, a lawyer is unlikely to agree to represent a case unless there is a potential for return, most typically when the case is strong and there is some indication of liability (Eisenberg & Lanvers, 2009). Without a contingent fee structure, plaintiffs who are generally "lower to middle income" employees (Colvin, 2007) must assume the costs themselves. As such, at trial employees are most often represented by themselves or inexperienced (and thus more affordable) counsel (Estreicher et al., 2017).

Clearly, pursuing a claim can be a more expensive endeavor for an employee with fewer resources than it is for an organization. Donohue and Siegelman (1991) suggested that it is economically unrealistic for plaintiffs to pursue cases if they earn less than $450 a week. Adjusting for inflation, this estimate would be $864 today. According to the Bureau of Labor Statistics, the median weekly earnings of full-time wage and salary workers was $957 (Bureau of Labor Statistics, 2020). The median weekly earnings for employees who are Black ($775), Hispanic ($722), or female ($857) were lower and would not reach the $864 threshold. This estimate conveys that many already disadvantaged employees are unlikely to even consider pursuing litigation and if they did, they would experience a considerable economic hardship.

6.3.2. ***Decisions for the Employee***

There were 15 cases in this sample where an employee was ultimately successful in defending their claims. Of these cases 12 brought claims under Title VII, 8 included a claim of Title VII retaliation, there were 3 claims under Section 1983, 2 contained ADEA claims, and 1 ADA claim. In some of these cases the plaintiffs had successfully defended at least one of their claims. In the case of *Davis v. Florida. Agency for Health Care Administration* (2015), the employee successfully defended their Title VII retaliation claim and was awarded compensatory damages and did so again when the organization appealed the trial decision.

The PAs presented as evidence in this subset of cases were often used to demonstrate the employee's satisfactory performance. Evidence of satisfactory performance from PAs was found in all but one of the successful cases. In an ADEA case, an employee used their positive PA ratings to demonstrate that they were performing satisfactorily prior to elimination and the immediate replacement by a younger applicant could not have been performance-related (*Pierson v. Quad/Graphics Printing Corporation, 2016*). The PAs also appeared to be used to demonstrate trends in the employee's performance history. In *Hague v. University of Texas Health Science Center*, (2014) the performance-related reasons for not renewing an employee's contract did not match the satisfactory performance reflected in the employee's appraisals.

Although rare, organizations can also be caught altering or falsifying performance information and PAs. This occurred in several of the cases with decisions for the employee. In one case (*Goudeau v. Nat'l Oilwell Varco, L.P.,* 2015) four time-stamped written warnings were issued to an employee after the employee was terminated. These warnings were untimely and utilized performance incidents across the employee's several year career. In an ADA case (*Burton v. Freescale Semiconductor, Inc.*, 2015), an employee presented evidence that they were

terminated shortly after disclosing health problems and revealed that forged PAs were created only after the termination. This forged PA evidence was used to argue that termination based on poor performance was pretext for disability discrimination.

While not explicit recommendations, using PAs as a retaliatory tool, inaccurately reflecting performance, and altering or forging PAs are clearly terrible breaches of defensible PA procedures. The relationship between declines in PA ratings following a protected action and decisions for employees was not quantitatively supported in hypothesis testing. However, the presence of this pattern in the handful of cases found in favor of the employee, suggest it is an important phenomenon. These cases support the importance of temporal measurement and documentation of performance in PA for proving a retaliatory timeline before and after a protected action. Additionally, PAs can be used as evidence of pretextual intent to discriminate against an individual. In the examples provided above, each of the PAs were used as justification for why the employee was terminated or their contract was not renewed. These justifications were considered to be an attempt to cover up behaviors that knew were illegal.

6.3.3. ***Performance Improvement Opportunities***

When organizational PA procedures note deficiencies and provide employees with an opportunity to improve their performance, litigation decisions were significantly more likely to favor the organization than the employee. This result was gathered from the combination of both structured and unstructured opportunities. The presence of evidence for these opportunities was related to the greatest increase in the odds of a decision for the organization compared to all other significantly related recommendations. For organizations this result should be compelling and the importance of following this recommendation should be highlighted. Given the potential

importance of this relationship and many ways an organization can provide this opportunity, some elaboration on PIPs is warranted.

A PIP is a structured process that informs the employee of the performance expectations they are not meeting, clarifies expectations, and provides guidance to facilitate improvement. Unfortunately, there is very little information about PIPs in the empirical literature. However, PIPs can be considered a part of a greater PM process and share many of the same best practice recommendations as PAs. Like PA, the expectations set in a PIP should be understood by the employee and they should be provided with a reasonable timeline to make their improvements. These recommendations are important and critical to an employee's perception of the fairness (Greenburg, 1986; Gruman & Saks, 2011; Stiles, Gratton, Truss, Hope-Hailey, & McGovern, 1997; Tanner, 2012). Ideally, they are also accompanied by more regular meetings where progress is appraised. In these meetings, the rater has the opportunity to document any improvements or issues as they occur (Suttapong, Srimai, & Pitchayadol, 2013). As the end of the PIP nears, the rater should determine if the employee met the expectations laid out in the PIP (Tanner, 2012). Other areas that might need improvement that come up along the way or that are not essential to the functions of the job cannot sway the decision. Collaboration amongst the Human Resources department, the manager and management should take place to determine how to move forward at the end of the PIP time period (SHRM, 2015). If expectations are met, the employee should be relieved of the added monitoring and released to perform their job normally. If not met, the consequences agreed upon should be carried out.

PIP opportunities were coded separately from informal opportunities to improve, because they represent a unique formal extension of the PA process. Of the 83 cases with PIP evidence, only five cases resulted in decisions favoring the employee. In the majority of cases

organizations were attempting to defend performance-related reasons for making the employment decisions that the plaintiff alleged were adverse. Many cases with explicit mention of a formal PIP opportunity for improvement also included detailed timelines following the meetings and PAs within the PIP period. The presence of this amount of detail was unique to formal opportunities to improve and the attention placed on how the employee's PIP progressed demonstrates the importance that these judges placed on this practice.

6.4. Limitations

6.4.1. *Representation of Discrimination Claims*

Discrimination cases are greatly underrepresented in employment litigation due to the large number of ways a claim can be dropped or settled before reaching a courtroom (Clermont & Schwab, 2004). From the time that an employee identifies or is subjected to some discriminatory treatment to receiving the "right to sue" notice from the EEOC, there are a number of increasingly difficult hurdles that the employee must face and overcome. At any point in this initial process an employee may feel discouraged and cease the pursuit of their claims, they could miss a critical reporting deadline and be time-barred from pursuing litigation, they could fail to find legal counsel or their legal counsel could convince them that their case is unlikely to succeed. Organizations may also seek to resolve the claims prior to public trial. Cases like those examined in this review where an employee has pursued a claim to the level of trial and subsequently appeal that trial decision are even rarer and are likely not representative of the universe of employment law claims.

Arbitration and the practice of settling claims outside of a courtroom have contributed to the phenomena of the "vanishing trial" (Galanter, 2004). Arbitration is a form of alternative dispute resolution and is a method to resolve disputes outside the courts. Cases that reach trial

are a minority and settlement decisions account for a great majority of case resolutions (Estreicher et al., 2017). Using arbitration to reach settlement is especially prevalent in employment litigation (Colvin & Pike, 2014). There is some research that suggests employees, organizations, and even judges prefer arbitration to trial (Resnik, 1994, 2002). In a study of 7,316 cases from 2012-2017 arbitrated by the American Arbitration Association, 75.7% of employment cases were settled prior to a trial (Estreicher et al., 2017). Some organizations have attempted to utilize internal arbitration before a complaint is escalated to a formal claim with the EEOC (Van Loo, 2016). Settlements often involve confidentiality agreements that creating an unobservable population of cases characterized by the moniker of "invisible settlements" (Gross & Syverud, 1996; Resnik, 2006).

According to some scholars, the large number of invisible settlements has made discrimination in the workplace almost impossible to investigate (Brinn, 2019; Kotkin, 2005). In addition to losing information that would be publicly available if these cases reached trial, settlements often involve confidentiality agreements. The resolution records of invisible settlements only include a dismissal that does not identify the terms of agreement and confidentiality clauses prevent plaintiffs from disclosing payments or, in some cases, that a claim was ever made (Kotkin 2005; Weston, 2020).

6.4.2. ***Judicial Bias***

There may be a perception by some people that employment discrimination cases are easy to win (Selmi, 2000). Selmi (2001) examined plaintiff success rates in employment trials and found that plaintiffs were more likely to gain favorable decisions in a jury-trial (39.9%) compared to non-jury trials (18.7%). This was a notable difference compared to insurance and personal injury trials where there was not a marked disparity between jury and non-jury

decisions (Selmi, 2001). These findings have been attributed to judicial bias against discrimination cases (Selmi, 2001; see also Cude & Steger, 1998 and Clermont & Schwab, 2004). Oppenheimer (2003) found that, perhaps unsurprisingly, this bias disproportionately disadvantages women and minorities. Indeed, further legal empirical examinations have supported these findings and demonstrate that plaintiffs in employee discrimination suits fare worse than other civil plaintiffs (see Kotkin, 2005).

6.4.3. *Case Decisions*

The inclusion of case decisions that have not been finalized introduces a level of uncertainty to the results of the analysis because the final decision once at trial is unknown. Decisions in favor of the employee were counted alongside any appeal decisions that included a reversal of a plaintiffs claim. Any number of actions can follow an appellate reversal. For example, organizations and employees can settle outside of court and a reversed decision will not be seen at trial again. Also, cases can span the course of several years. Some cases in this sample that were appealed in 2014 still have not been fully resolved almost six years later.

6.5. Future Directions

Examining litigation using this method requires a considerable investment of time and effort. Reading and extracting valuable information in narratives that often summarize an individual's entire employment history is difficult because of the sheer amount of information in these narratives. Even focusing solely on PA resulted in a large amount of information that could be extracted. Because this information is not presented uniformly, or at all in some cases, this examination is difficult. Future investigations could take a more targeted approach to case selection. Narrowing the sample of cases and conducting a more in-depth treatment of one

specific legal protection (e.g., ADA, ADEA, or Section 1981 claims) could reveal what elements or characteristics of PAs are most relevant to these particular legal protections.

A great deal can also be learned from examining claims where the employee is victorious. These 15 cases where decisions were in favor of the employee contained more PA information than the average case examined and coded. Judges appeared to spend more time describing the history of the employee's performance. Perhaps when claims are stronger and result in a victory more attention is paid to documenting the details of the case in its decision. Written opinions of cases that failed on a procedural legal ground (e.g., missing a reporting deadline) tended to be the shortest, more focused on the letter of the law, and least fact-intensive.

This review considered a large number of variables related to PA and focusing the analysis on a subset of these variables could have made the review more manageable. In particular, an examination of how the features and practices of opportunities to improve fairness in litigation is theoretically interesting due to the strength of the observed relationship to case decisions. Cases with PIPs and formal opportunities to improve performance contained more performance-related detail and often followed the process prior to the PIP initiation to its conclusion. The opportunity to improve proved to be important to the odds of an organization's victory and future examinations of discrimination claims might identify specific ways organizations give employees opportunities to develop.

The desire to eliminate PAs from practice raises the question of how organizations will legally defend personnel decisions. Whereas this study does not systematically compare organizations that do and do not conduct PAs, the results do raise doubts about how successful these organizations would have been if they did not have PA evidence.

Future research is needed to examine whether PAs are the source of the issues that many have attributed to PA or a sign of a larger issue. It appears that organizations and individuals that advocate for the abandonment of PAs and ratings believe the faults of PA are due to the PA itself. Perhaps the source of the issues identified with PA are the characteristics and practices built into the design of the PA procedure (Gorman et al., 2017). Prior to making these decisions the organization must review how they appraise employee performance and examine whether they adhere to best practice recommendations. The effectiveness of a PA is bound to the contextual and situational constraints and demands it operates under. An internally-focused examination should be conducted prior to system overhauls. Further, whether the PA in and of itself is the problem has yet to be proven empirically and it is left to future examinations to determine.

PAs serve critical purposes; organizations need to evaluate their employee's effectiveness and make informed administrative decisions while employees need to be made aware of how their performance fares relative to expectations and be given the opportunity to improve. To this point most organizations have used one PA to address all of these needs. The literature suggests that rather than focusing centrally on the traditional-depicted PA, organizations should adopt a systems approach (Schleicher et al. 2019) that can serve and resolve each of these purposes independently while providing a richer source of information for the organization and supplementing the growth of the employee. Research should examine the effectiveness of a systems approach.

7. CONCLUSION

Reading hundreds of appellate cases where employees made claims against their organizations illuminated the wealth of information present within case narratives and how further mining of this source can provide critical and important information. While it was assumed that case narratives would contain both PA content and process-related information, in this examination there was considerably more process-related information than content-related information. Ultimately, the use of multiple raters, consistent and thorough PA documentation, allowing employees to review PA results, improvement opportunities, and following more of Malos's recommendations were related to decisions favoring the organization. Additionally, an organization's sector and the presence of satisfactory PA evidence also revealed relationships to the likelihood of case decisions. Suggestions for how these findings can be utilized follows.

Organizations may be interested to discover that in this examination the content and criteria within the PA itself were not present in the case narratives. Organizations should still follow recommendations for the appropriate development and selection of PA criteria. This does not mean valid PAs are not necessary or that the process of developing the PA criteria should be neglected. Rather it highlights the relative importance of the PA process. The evidence presented in employment litigation centers around the employee's experiences and as such details about the procedure followed when conducting the employee's PAs are likely to appear when PAs are used as evidence in discrimination cases. Care should be taken to ensure the PA process is fair to employees.

Fairness and perceptions of fairness remain important to the success and legal defensibility of PA process recommendations. Organizations should ensure that employees are

aware of how their current performance is perceived relative to what is expected of them. Organizations and supervisors should prioritize the provision of improvement opportunities to underperforming employees. This practice allows the employee to address their performance deficiencies. Tthe success or failure of these opportunities can clarify why the administrative decision was made.

Organizations that are interested in maximizing the legal defensibility of their PAs must consider how their system trains raters and the ways the PA is used. Organizations did not fare well when supervisors used PAs as retaliatory tools or as a means of pretext. This usage is unfair to the employee and often illegal if following a protected activity or if it results in discriminatory action. Organizations can pre-empt the potential use of PAs for these disreputable purposes by clearly communicating the purpose of the PA and imposing consequences for their misuse on supervisors. Organizations could also ensure that if there is a system for employees to report abuse and misuse of PA, that there are no barriers that would prohibit or discourage employees from using it. When these systems do not exist, they should be created and organizations should communicate their availability to employees.

No PA system is perfect. Even if it were, a perfect system cannot make up for the errors that may occur in the process of utilizing of the PA system. PAs often serve as a source of evidence in employment discrimination litigation. The information present in the PA can be used by and benefit either the organization or the employee. Given this relationship and the results of this examination, organizations should follow the PA recommendations considered in this review and generated from the PA literature. While an organization's PA system will never be perfect, it can be legally defensible.

REFERENCES

Adler, S., Campion, M., Colquitt, A., Grubb, A., Murphy, K., Ollander-Krane, R., & Pulakos, E. D. (2016). Getting rid of performance ratings: Genius or folly? A debate. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *9*, 219-252. https://doi.org/10.1017/iop.2015.106

Age Discrimination in Employment Act, 29 U.S.C. § 621 (1967).

Aguinis, H. (2013). *Performance management* (3rd ed). Upper Saddle River, NJ: Pearson/Prentice Hall.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2012). Delivering effective performance feedback: The strengths-based approach. *Business Horizons, 55,* 105-111. https://doi.org/10.1016/j.bushor.2011.10.004

Aguinis, H., Joo, H., & Gottfredson, R. K. (2011). Why we hate performance management—And why we should love it. *Business Horizons, 54,* 503-507. https://doi.org/10.1016/j.bushor.2011.06.001

Americans with Disabilities Act of 1990, Pub. L. No. 101-336, 104 Stat. 328 (1990).

Ashe Jr, R. L., & McRae, G. S. (1984). Performance evaluations go to court in the 1980's. *Mercer Law Review, 36*, 887-905.

Barrett, G. V., & Kernan, M. C. (1987). Performance appraisal and terminations: A review of court decisions since Brito v. Zia with implications for personnel practices. *Personnel Psychology*, *40*, 489-503. https://doi.org/10.1111/j.1744-6570.1987.tb00611.x

Basnight, T. A., & Wolkinson, B. W. (1977). Evaluating managerial performance: Is your appraisal system legal? *Employee Relations Law Journal, 3*, 240-254.

Beck-Dudley, C. L., & McEvoy, G. M. (1991). Performance appraisals and discrimination suits: do courts pay attention to validity? *Employee Responsibilities and Rights Journal, 4*, 149-163. https://doi.org/10.1007/BF01390356

Becker, T. E. & Cote, J. A. (1994). Additive and multiplicative method effects in applied psychological research: An empirical assessment of three models. *Journal of Management, 20*, 625-641. https://doi.org/10.1016/0149-2063(94)90006-X

Becker, T. E., & Klimoski, R. J. (1989). A field study of the relationship between the organizational feedback environment and performance. *Personnel Psychology*, *42*, 343-358. https://doi.org/10.1111/j.1744-6570.1989.tb00662.x

Bernardin, H. J., Kane, J. S., Ross, S., Spina, J. D., & Johnson, D. L. (1995). Performance appraisal design, development, and implementation. In G.R. Ferris, S, D, Rosen & D. T. Barnum (Eds.), *Handbook of human resource managemen*t (pp. 462-493).Cambridge, MA: Blackwell.

Buckingham, M., & Goodall, A. (2015, April). Reinventing performance management. *Harvard Business Review*. https://hbr.org/2015/04/reinventing-performance-management

Boswell, W. R., & Boudreau, J. W. (2002). Separating the developmental and evaluative performance appraisal uses. *Journal of Business and Psychology*, *16*, 391-412. https://doi.org/10.1023/A:1012872907525

Boyd, C. L., Epstein, L., & Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science*, *54*, 389-411.

Burchett, S. R., & De Meuse, K. P. (1985). Performance appraisal and the law. *Personnel*, *62*, 29-37.

Burke, R. J., & Wilcox, D. S. (1969). Characteristics of effective employee performance review and development interviews. *Personnel Psychology*, *22*, 291-305. https://doi.org/10.1111/j.1744-6570.1978.tb02130.x

Carp, R. A., Manning, K. L., Holmes, L. M., & Stidham, R. (2019). *Judicial process in America*. CQ Press.

Cascio, W. F., & Aguinis, H. (2014). *Applied psychology in human resource management*. New York, NY: Pearson Higher Ed.

Cascio, W. F., & Bernardin, H. J. (1981). Implications of performance appraisal litigation for personnel decisions. *Personnel Psychology*, *34*, 211-226. https://doi.org/10.1111/j.1744-6570.1981.tb00939.x

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, *83*, 615-633. https://doi.org/10.1037/0021-9010.83.4.615

CBOCS West, Inc. v. Humphries, 553 U.S. 442, 128 S. Ct. 1951, 170 L. Ed. 2d 864 (2008).

Civil Rights Act of 1964, Pub.L. 88-352, 78 Stat. 241 (1964).

Clermont, K. M., & Schwab, S. J. (2004). How employment discrimination plaintiffs fare in federal court. *Journal of Empirical Legal Studies*, *1*, 429-458.

Cleveland, J. N. & Murphy, K. R. (2016). Organizations want to abandon performance appraisal: Can they? Should they? In D. L. Stone & J. H. Dulebohn (Eds). *Human resource management theory and research on new employment relationships* (pp. 15-46). Charlotte, NC: Information Age.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, *74*, 130-135. https://doi.org/10.1037/0021-9010.74.1.130

Coens, T., & Jenkins, M. (2000). *Abolishing performance appraisals: Why they backfire and what to do instead*. Oakland, CA: Berrett-Koehler.

Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *86*, 278-321. https://doi.org/10.1006/obhd.2001.2958

Colvin, A. J. (2007). Empirical research on employment arbitration: Clarity amidst the sound and fury. *Employee Rights and Employment Policy Journal*, *11*, 405-447.

Colvin, A. J., & Pike, K. (2014). Saturns and richshaws revisited: What kind of employment arbitration system has developed. *Ohio State Journal on Dispute Resolution*, *29*, 59-84.

Cude, D. S., & Steger, B. M. (1998). Does justice need glasses-unlawful retaliation under Title VII following Mattern: Will courts know it when they see it? *The Labor Lawyer, 14*, 373-413.

Davidson, J. E. (1995). The temptation of performance appraisal abuse in employment litigation. *Virginia Law Review*, *81*, 1614-1617. 10.2307/1073533 https://www.jstor.org/stable/i244392

De Mesquita, E. B., & Stephenson, M. (2002). Informative precedent and intrajudicial communication. *American Political Science Review, 96*, 755-766. https://doi.org/10.1017/S0003055402000436

DeNisi, A. S., & Pritchard, R. D. (2006). Performance appraisal, performance management, and improving individual performance: A motivational framework. *Management and Organization Review, 2*, 253–277. https://doi.org/10.1111/j.1740-8784.2006.00042.x

Dipboye, R. L., & De Pontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. *Journal of Applied Psychology*, *66*, 248-251. https://doi.org/10.1037/0021-9010.66.2.248

Donohue III, J. J., & Siegelman, P. (1991). The changing nature of employment discrimination litigation. *Stanford Law Review*, 983-1033.

EEOC v. Ford Motor Co., 782 F.3d 753 (6th Cir. 2015).

Eisenberg, T., & Lanvers, C. (2009). What is the settlement rate and why should we care? *Journal of Empirical Legal Studies*, 111-146.

El Kharzazi, D. J., Siwatu, M., & Brooks, D. R. (2014). *Retaliation - making it personal*. https://www.eeoc.gov/laws/types/retaliation_considerations.cfm

Estreicher, S., Heise, M., & Sherwyn, D. S. (2017). Evaluating employment arbitration: A call for better empirical research. *Rutgers University Law Review.*, *70*, 320-347.

Estreicher, S., Heise, M., & Sherwyn, D. S. (2018). Evaluating employment arbitration: Call for better empirical research. *Rutgers University Law Review, 70*, 375-400.

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, 25,* 392-406. https://doi.org/10.5465/255999

Feild, H. S., & Thompson, D. T. (1984). Study of court decisions in cases involving employee performance appraisal systems. *Daily Labor Report* (BNA*), 248*, E-1.

Fox v. Leland Volunteer Fire/Rescue Department, Inc., No. 15-1364 (4th Cir. May 5, 2016).

Galanter, M. (2004). The vanishing trial: An examination of trials and related matters in federal and state courts. *Journal of Empirical Legal Studies*, *1*, 459-570.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management*, *16*, 399-432. https://doi.org/10.1177/014920639001600208

Greenberg, J., & Colquitt,J. (2005). *Handbook of organizational justice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Godwin v. Wellstar Health System, Inc., No. 14-11637 (11th Cir. June 17, 2015).

Goldman, B. M. (2001). Toward an understanding of employment discrimination claiming: An integration of organizational justice and social information processing theories. *Personnel Psychology*, *54*, 361-386. https://doi.org/10.1111/j.1744-6570.2001.tb00096.x

Gross, S. R., & Syverud, K. D. (1996). Don't try: Civil jury verdicts in a system geared to settlement. *University of California Los Angeles Law Review.*, *44,* 1-64.

Guthrie, C., & George, T. E. (2004). The futility of appeal: Disciplinary insights into the affirmance effect on the United States Courts of Appeals. *Florida State University Law Review*, *32*, 357-385. http://ssrn.com/abstract=743167

Henry v. Federal Reserve Bank of Atlanta, No. 14-5618 (6th Cir. Apr. 14, 2015).

Hettal, K. A., & Garza, M. (Co-Chairs) (2016, April). *Where have all the ratings gone?* Interactive discussion at the Society for Industrial and Organizational Psychology annual conference. Anaheim, CA.

Huber, V. L., Neale, M. A., & Nofthcraft, G. B. (1987). Judgment by heuristics: Effects of ratee and rater characteristics and performance standards on performance-related judgments. *Organizational Behavior and Human Decision Processes*, *40*, 149-169. https://doi.org/10.1016/0749-5978(87)90010-0

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, *50*, 905-925. https://doi.org/10.1111/j.1744-6570.1997.tb01487.x

Kacmar, K. M., Witt, L. A., Zivnuska, S., & Gully, S. M. (2003). The interactive effect of leader-member exchange and communication frequency on performance ratings. *Journal of Applied Psychology*, *88*, 764-772. https://psycnet.apa.org/doi/10.1037/0021-9010.88.4.764

Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reaction: Measurement, modeling, and method bias. *Journal of Applied Psychology, 85,* 708-723. https://doi.org/10.1037/0021-9010.85.5.708

Kleiman, L. S., & Durham, R. L. (1981). Performance appraisal, promotion and the courts: A critical review. *Personnel Psychology*, *34*, 103-121. https://doi.org/10.1111/j.1744-6570.1981.tb02181.x

Kotkin, M. J. (2005). Invisible settlements, invisible discrimination. *North Carolina Law Review*, *84*, 927-978.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology*, *70*, 56-65. https://doi.org/10.1037/0021-9010.70.1.56

Kuvas, B. (2006). Performance appraisal satisfaction and employee outcomes: Mediating and moderating roles of work motivation. *International Journal of Human Resource Management*, 17, 504–522. https://doi.org/10.1080/09585190500521581

Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: Challenging the “more is better”

assumption. *Organizational Behavior and Human Decision Processes*, *116*, 217-228. https://doi.org/10.1016/j.obhdp.2011.05.002

Latham, G. P., & Wexley, K. N. (1980) *Increasing productivity through performance appraisal.* New York: Addison-Wesley.

Leventhal, G. S. (1980). What should be done with equity theory? In *Social exchange* (pp. 27-55). Boston, MA: Springer.

Levy, P. E., Tseng, S. T., Rosen, C. C., & Lueke, S. B. 2017. Performance management: A marriage between practice and science: Just say "I do." In M. R. Buckley, A. R. Wheeler, & J. R. B. Halbesleben (Eds.), *Research in personnel and human resources management*, (*35*, pp.155-213). Bingley, England: Emerald.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Boston, MA: Springer Science & Business Media. https://doi.org/10.1007/978-1-4899-2115-4_8

Lucas v. WW Grainger, Inc., 257 F.3d 1249 (11th Cir. 2001).

Lubben, G. L., Thompson, D. E., & Klasson, C. R. (1980). Performance appraisal: the legal implications of Title VII. *Personnel*, *57*, 11-21.

Malos, S. B. (1998). Current legal issues in performance appraisal. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 49–94). San Francisco, CA: Jossey-Bass.

Malos, S. B. (2005). The importance of valid selection and performance appraisal: Do management practices figure in case law? In F. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 373-409). San Francisco, CA: Jossey-Bass.

Martin, D. C., & Bartol, K. M. (1991). The legal ramifications of performance appraisal- an update. *Employee Relations Law Journal*, *17*, 257-286.

Martin, D. C., Bartol, K. M., & Kehoe, P. E. (2000). The legal ramifications of performance appraisal: The growing significance. *Public Personnel Management*, *29*, 379-406. https://doi.org/10.1177/009102600002900307

Martin, D. C., Bartol, K. M., & Levine, M. J. (1986). The legal ramifications of performance appraisal. *Employee Relations Law Journal*, *12*, 370-395.

Martin, R., Guillaume, Y., Thomas, G., Lee, A., & Epitropaki, O. (2016). Leader–member exchange (LMX) and performance: A meta-analytic review. *Personnel Psychology*, *69*, 67-121. https://doi.org/10.1111/peps.12100

Martin v. Eli Lilly & Co., No. 16-11537 (11th Cir. July 21, 2017).

Meyer, H. H. (1980). Self-appraisal of job performance. *Personnel Psychology*, *33*, 291-295. https://doi.org/10.1111/j.1744-6570.1980.tb02351.x

Meyer, H. H., & Walker, W. B. (1961). A study of factors relating to the effectiveness of a performance appraisal program. *Personnel Psychology*, *14*, 291-298. https://doi.org/10.1111/j.1744-6570.1961.tb01235.x

Nagelkerke, N. J. D. 1991. A note on the general definition of the coefficient of determination. *Biometrika, 78,* 691-692. https://www.jstor.org/stable/2337038

Nickols, F. (2007). Performance appraisal: Weighed and found wanting in the balance. *The Journal for Quality and Participation*, *30*, 13-16.

Norris-Watts, C., & Levy, P. E. (2004). The mediating role of affective commitment in the relation of the feedback environment to work outcomes. *Journal of Vocational Behavior*, *65*, 351-365. https://doi.org/10.1016/j.jvb.2003.08.003

Oppenheimer, D. B. (2003). Verdicts matter: An empirical study of California employment discrimination and wrongful discharge jury verdicts reveals low success rates for women and minorities. *University of California Davis Law Review*, *37*, 511-567.

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology*, *78*, 569-582. https://doi.org/10.1037/0021-9010.78.4.569

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment and memory* (79-96). Hove, NY : Psychology Press.

Practical Law Labor and Employment. (2013). *That's so 1981: Comparing Title VII and Section 1981*.https://content.next.westlaw.com/Document/Ia7aac9dc642211e38578f7ccc38dcbee/View/FullText.html?contextData=(sc.Default)&transitionType=Default&firstPage=true&bhcp=1

Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology*, *74*, 770-780. https://doi.org/10.1037/0021-9010.74.5.770

Resnik, J. (1993). Whose judgment-vacating judgments, preferences for settlement, and the role of adjudication at the close of the twentieth century. *University of California in Los Angeles Law Review*, *41*, 1471-1539.

Resnik, J. (2002). Mediating preferences: Litigant preferences for process and judicial preferences for settlement. *Journal of Dispute Resolution*, *1*, 155-169.

Resnik, J. (2006). Uncovering, disclosing, and discovering how the public dimensions of court-based processes are at risk. *Chicago-Kent Law Review, 81*, 521-570.

Roberts, G. E. (2003). Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, *32*, 89-98. https://doi.org/10.1177/009102600303200105

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, *85*, 370-395. https://doi.org/10.1111/j.2044-8325.2011.02045.x

Rock, D., Davis, J., & Jones, B. (2014, August 8). *Kill your performance ratings. Strategy + Business.* https://www.strategy-business.com/article/00275

Rock, D., & Jones, B. (2016, September). Why more and more companies are ditching performance ratings. *Harvard Business Review*. https://hbr.org/2015/09/why-more-and-more-companies-are-ditching-performance-ratings.

Rynes, S. L., Gerhart, B., & Parks, L. (2005). Personnel psychology: Performance evaluation and pay for performance. *Annual Review of Psychology, 56*, 571-600. https://doi.org/10.1146/annurev.psych.56.091103.070254

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, *95*, 355-386. https://doi.org/10.1037/0033-2909.95.3.355

Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, *78*, 450-462. https://doi.org/10.1037/0021-9010.78.3.450

Schleicher, D. J., Baumann, H. M., Sullivan, D. W., Levy, P. E., Hargrove, D. C., & Barros-Rivera, B. A. (2018). Putting the system into performance management systems: A

review and agenda for performance management research. *Journal of Management*, *44*, 2209-2245. 10.1177/0149206318755303

Selmi, M. (2000). Why are employment discrimination cases so hard to win. *Louisianna Law Review*, *61*, 555-575.

Stauffer, J. M., & Buckley, M. R. (2005). The existence and nature of racial bias in supervisory ratings. *Journal of Applied Psychology*, *90*, 586 –591. https://doi.org/10.1037/0021-9010.90.3.586

Steelman, L. A., Levy, P. E., & Snell, A. F. (2004). The feedback environment scale: Construct definition, measurement, and validation. *Educational and Psychological Measurement*, *64*, 165-184. https://doi.org/10.1177/0013164403258440

Tabachnick, B. G., & Fidel, L. S. (2001). *Using multivariate statistics* (4th ed.). Needleham Heights, MA: Allyn & Bacon.

Taylor, E. K., & Wherry, R. J. (1951). A study of leniency in two rating systems. *Personnel Psychology*, *4*, 39-47. https://doi.org/10.1111/j.1744-6570.1951.tb01459.x

Terpstra, D. E., & Honorée, A. L. (2016). Differences in the nature of employment discrimination litigation between private sector organizations and public sector organizations. *International Journal of Discrimination and the Law*, *16*, 200-213. https://doi.org/10.1177/1358229116645677

The Pregnancy Discrimination Act, 42 U.S.C. § 701 (1978).

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.

Thornton, G. C., & Zorich, S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology*, *65*, 351-354. https://doi.org/10.1037/0021-9010.65.3.351

Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C., & Privette, M. (2008). Anchoring effects on performance judgments. *Organizational Behavior and Human Decision Processes*, *107*, 29-40. https://doi.org/10.1016/j.obhdp.2008.01.003

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131. 10.1126/science.185.4157.1124

U.S. Equal Employment Opportunity Commission. (n.d.). *Charge statistics (charges filed with EEOC) FY 1997 through FY 2019*. EEOC Enforcement & Litigation Statistics. https://www.eeoc.gov/eeoc/statistics/enforcement/charges.cfm

U.S. Employment Opportunity Commission. (n.d.b). *Filing a lawsuit*. https://www.eeoc.gov/employees/lawsuit.cfm

U.S. Equal Employment Opportunity Commission. (n.d.c). *Litigation statistics FY 1997 through FY 2019*. EEOC Enforcement & Litigation Statistics. https://www.eeoc.gov/eeoc/statistics/enforcement/litigation.cfmU.S. Equal

Van Loo, R. (2016). The corporation as courthouse. *Yale Journal on Regulation*, *33*, 547-602.

Veglahn, P. A. (1993). Key issues in performance appraisal challenges: Evidence from court and arbitration decisions. *Labor Law Journal, 44*, 595-606.

Wells, R. G. (1982). Guidelines for effective and defensible performance appraisal systems. *Personnel Journal*, *61*, 776-782.

Werbel, J., & Balkin, D. B. (2010). Are human resource practices linked to employee misconduct? A rational choice perspective. *Human Resource Management Review*, *20*, 317-326. https://doi.org/10.1016/j.hrmr.2009.10.002

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. https://doi.org/10.1111/j.1744-6570.1997.tb00898.x

Weston, M. (2020). Buying secrecy: Non-Disclosure agreements, arbitration, and professional ethics in the # MeToo Era. In *Arbitration, and Professional Ethics in the# MeToo Era*, *University of Illinois Law Review, 5*. http://dx.doi.org/10.2139/ssrn.3542590

Whitaker, B. G., Dahling, J. J., & Levy, P. (2007). The development of a feedback environment and role clarity model of job performance. *Journal of Management*, *33*, 570-591. https://doi.org/10.1177/0149206306297581

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. https://doi.org/10.1007/s10869-013-9299-4

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*, 189–205. https://doi.org/10.1111/j.2044-8325.1994.tb00562.x

APPENDIX A

Table A

Summarized Cases with Decisions in Favor of the Employee

Case Name [a]	Successful Claims	Appeal Summary [b]	Presence of Performance Appraisal
Bonenberger v. St. Louis Metro. Police Dep't	Race discrimination claims under Title VII, 42 U.S.C.S. §§ 1983 and 1981	Where a jury found in employee's favor on his race discrimination claims under Title VII, 42 U.S.C.S. §§ 1983 and 1981, against three superiors, the superiors' motion for judgment as a matter of law was properly denied because the position offered a material change in working conditions.	Employee alleged that he applied for a promotion and was denied the position because of his race. Positive performance appraisals are used as evidence to demonstrate his qualifications relative to other applicants.
Davis v. Fla. Agency for Health Care Admin.	Title VII Retaliation	Employee presented sufficient evidence for reasonable jury to conclude that she suffered retaliation under Title VII and Florida Civil Rights Act because, inter alia, she presented enough evidence for jury to reasonably conclude that her grievance was the but-for cause of the retaliatory action; evidence supported award of compensatory damages.	Dated performance appraisals demonstrated a noted decrease in performance following her protected activity.

Table A Continued

Case Name [a]	Successful Claims	Appeal Summary [b]	Presence of Performance Appraisal
Hague v. Univ. of Tex. Health Sci. Ctr.	Title VII Retaliation	Employee's Title VII retaliation claim survived because she demonstrated pretext since, inter alia, two other female employees who supported her complaint were terminated, and evidence showed conflict regarding department head's stated reasons for not renewing her contract.	Performance-related reasons for not renewing her contract did not match the satisfactory performance reflected in the employee's appraisals.
Pierson v. Quad/Graphics Printing Corp.	Age Discrimination in Employment Act	Terminated employee's age discrimination claims under ADEA survived because genuine factual dispute existed regarding whether his position was eliminated in reduction in force or whether he was replaced by younger individual; he presented sufficient evidence of pretext in the form of shifting justifications.	Positive performance appraisals presented as evidence that the employee was performing satisfactorily prior to elimination and replacement by a younger applicant.
Yazdian v. ConMed Endoscopic Techs., Inc.	Title VII Retaliation	Employee was terminated for insubordination and unprofessional behavior eight weeks after he accused his supervisor of creating a hostile work environment, his Title VII retaliation claim survived .	After a history of positive appraisals, the employee received a negative appraisal following a protected activity. Within the negative performance appraisal, the supervisor specifically referenced the employee's protected statements as examples of insubordination.

Table A Continued

Case Name [a]	Successful Claims	Appeal Summary [b]	Presence of Performance Appraisal
Crawford v. Duke	Title VII Race Discrimination	In this Title VII action, the dismissal for failure to exhaust administrative remedies was reversed in part because the employee adequately exhausted his administrative remedies with respect to the October 2011 performance review and the December 2011 suspension, but not as to the November 2011 denial of promotion.	The performance review was previously used as evidence that the employee did not exhaust internal avenues with their employer. In appeal the employee demonstrated that he attempted to appeal the evaluation but nothing came of his discussions with his supervisor.
Goudeau v. Nat'l Oilwell Varco, L.P.	Age Discrimination in Employment Act	Former employee established a prima facie case of age discrimination under the ADEA; he also offered sufficient evidence of pretext given a supervisor's alleged ageist comments and the employee's claim that four written warnings relating to events on different dates were not given to him until the day he was fired.	The employer presented written warnings and performance motives as justification for the decision to terminate the employee. The employee contrasted a history of positive performance appraisals and revealed that the written performance warnings were only created the day of the termination.
Burton v. Freescale Semiconductor, Inc.	American's with Disability Act	Employee presented substantial evidence that termination based on poor performance was pretext for disability discrimination since reliance on dated performance reviews was dubious, supervisors provided conflicting testimony, exculpatory paper trail was created after decision to terminate, and termination occurred shortly after health problems.	Modified performance appraisals and previous performance appraisals from several years ago were used to defend an employee's termination. The employee was able to demonstrate that the negative appraisals were fabricated and did not accurately reflect their performance.

Table A Continued

Case Name [a]	Successful Claims	Appeal Summary [b]	Presence of Performance Appraisal
Miller v. Polaris Labs., LLC	Title VII Retaliation and Race Discrimination	Where employee was fired for inadequate production numbers, her race discrimination and retaliation claims under Title VII and 42 U.S.C.S. §§ 1981 and 1981a survived because there was evidence that her coworkers displayed discriminatory feelings toward her and that systematic tampering occurred to such an extent that it torpedoed her output.	Negative performance appraisals reflecting low production numbers were presented as evidence to support the employee's termination. The employee provided context to the negative appraisals by demonstrating a history of sabotage.
Zamora v. City of Houston	Title VII Retaliation	District court properly found employer liable on employee's Title VII of the Civil Rights Act of 1964 retaliation claim because employee produced evidence sufficient to find, under Nassar's but-for standard of causation, that his supervisors, motivated by retaliatory intent, intended to cause and did cause his suspension.	Employee performance appraisals reflected a history of satisfactory performance by supervisors.
Vega v. Hempstead Union Free Sch. Dist.	Title VII and § 1983 Race Discrimination	Employee sufficiently pleaded Title VII and § 1983 discrimination based on his allegation that the school district assigned him classes with higher numbers of Spanish-speaking students and, in doing so, assigned him a disproportionate workload.	Performance appraisals reflected positive performance prior to being assigned a larger workload. Negative performance appraisal following the new placement also references his assignment was a result of his race and ethnic background.

Table A Continued

Case Name [a]	Successful Claims	Appeal Summary [b]	Presence of Performance Appraisal
Brown v. Diversified Distrib. Sys., LLC	Title VII Sex Discrimination	Terminated employee's Title VII sex discrimination claim survived because she sufficiently demonstrated that her coworkers' actions were of "comparable seriousness" to the conduct for which she was discharged to establish a prima facie case, and circumstances would permit jury to infer that employer's justifications were pretextual.	Performance appraisals prior to termination demonstrated a history of positive performance. Employee who was fired for allegedly inappropriate work behavior presented performance appraisals of peers who performed similar activities and were not terminated as comparators.
Long v. Ala. Dep't of Human Res.	Title VII and § 1981 Retaliation	Where employee was terminated after he testified in coworker's discrimination suit, his retaliation claims under Title VII and § 1981 survived because, inter alia, a commissioner began the investigation that led to the employee's termination within a month of his first deposition, and just prior to his second deposition.	Performance appraisals demonstrated a history of positive performance. Negative performance documentation occurred only after the employee engaged in a protected activity.
Henry v. Abbott Labs.	Title VII Race Discrimination	Employee made out prima facie case of race discrimination under Title VII and Ohio Rev. Code Ann. § 4112.02(A) for 2009 and 2010 and could establish inference of discrimination, and genuine dispute of material fact existed about whether employer took materially adverse actions against her because of her protected activity.	Negative performance appraisal occurred only after the employee engaged in a protected activity. Prior performance appraisals demonstrated a history of positive performance.

Table A Continued

Case Name [a]	Successful Claims	Appeal Summary [b]	Presence of Performance Appraisal
Wilson v. Ark. Dep't of Human Servs.	Title VII Retaliation	Terminated employee's Title VII retaliation claim survived because the phrase "victim of retaliation, after having complained" alleged but-for causation, and six-week period between EEOC charge and the termination plausibly alleged a but-for causal connection; her disparate treatment claim failed.	Performance appraisals demonstrated a history of positive performance. Negative performance documentation occurred only after the employee engaged in a protected activity and included language that referenced the protected complaints.

Note. [a] Case names are presented with the abbreviations as they are given from NexisUni.
[b] Appeal summary is presented as it was given from NexisUni.

APPENDIX B

Table B1

Year the Appeal was Heard

Year	N Cases Examined	% of total
2014	103	22.3
2015	119	25.8
2016	68	14.7
2017	91	19.7
2018	81	17.5

Table B2

Court that Heard the Appeal

Appeals Court	N Cases Examined	% of total
United States Court of Appeals for the Federal Circuit	3	.6
United States Court of Appeals for the District of Columbia Circuit	10	2.2.
United States Court of Appeals for the First Circuit	15	3.2
United States Court of Appeals for the Second Circuit	36	7.8
United States Court of Appeals for the Third Circuit	46	10.0
United States Court of Appeals for the Fourth Circuit	22	4.8
United States Court of Appeals for the Fifth Circuit	50	10.8
United States Court of Appeals for the Sixth Circuit	82	17.7
United States Court of Appeals for the Seventh Circuit	52	11.3
United States Court of Appeals for the Eighth Circuit	68	14.7
United States Court of Appeals for the Ninth Circuit	17	3.7
United States Court of Appeals for the Tenth Circuit	34	7.4
United States Court of Appeals for the Eleventh Circuit	67	14.5

Table B3

State the Appeal was Originated From

State(s)[a]	Frequency (N) of Cases Examined	% of Total Cases
Michigan	38	8.2%
Illinois	36	7.8%
Pennsylvania	34	7.2%
Florida	33	7.1%
Texas	30	6.5%
New York	27	5.8%
Ohio	23	5%
Alabama	19	4.1%
Tennessee	17	3.7%
Washington; Georgia	16	3.5%
Colorado; Indiana; Maryland	12	2.6%
Mississippi	11	2.4%
Louisiana	10	2.2%
California	9	1.9%
Connecticut	8	1.7%
Arkansas; Iowa; Kentucky; Minnesota; Virginia	6	1.3%
Delaware; Massachusetts; New Jersey; Puerto Rico; Wisconsin	5	1.1%
Missouri; New Mexico; Nevada; Utah	4	0.9%
Kansas; North Carolina; Nebraska; Oklahoma	3	0.6%
Hawaii; Maine; South Dakota; Wyoming	2	0.4%
Arizona; Idaho; New Hampshire; South Carolina	1	0.2%

Note. [a]For the rows with multiple states, frequencies and percentages are for each state within that row.

APPENDIX C

Table C1

Correlations of all Case Decisions and Tested Variables

	Variable	1	2	3	4	5	6	7	8	9	10	11	12
	Hypothesis 6												
1.	More than One Rater	—											
2.	Number of Raters	.98*	—										
	Hypothesis 8												
3.	Number of Appraisals	.47*	.51*	—									
	Hypothesis 9_a												
4.	Opportunity to Improve	.20*	.19*	.40*	—								
	Hypothesis 10_a												
5.	Opportunity to Review	.07	.07	.41*	.48*	—							
	Hypothesis 11_a												
6.	(a) Ability to Appeal	.10	.13*	.29*	.28*	.38*	—						
7.	(b) Appraisal Appealed	-.12	-.15*	-.28*	-.24*	-.33*	-.86*	—					
	Hypothesis 13												
8.	Total Recs Followed	.50*	.50*	.72*	.64*	.76*	.60*	-.56*	—				
	Hypothesis 16_a												
9.	Public vs Private Sector	.03	.04	-.11*	-.09	.07	.05	-.06	.06	—			
	Hypothesis 15												
10.	Rating Decrease	.23*	.23*	.44*	.30*	.30*	.17*	-.16*	.41*	-.01	—		
	Hypothesis 17_a												
11.	Satisfactory Appraisals	.10	.10	.27*	.01	.21*	.11*	-.07	.27*	.03	.40*	—	
12.	Judicial Decisions	.05	.05	.08	.16*	-.02	-.02	.00	-.03	-.13*	-.05	-.18	—

*p <.05 (1-tailed).

Table C2

Correlations of Title VII Case Decisions and Tested Variables

	Variable	1	2	3	4	5	6	7	8	9	10	11	12
	Hypothesis 6												
1.	More than One Rater	—											
2.	Number of Raters	.97*	—										
	Hypothesis 8												
3.	Number of Appraisals	.47*	.51*	—									
	Hypothesis 9a												
4.	Opportunity to Improve	.20*	.19*	.40*	—								
	Hypothesis 10a												
5.	Opportunity to Review	.07	.07	.41*	.48*	—							
	Hypothesis 11a												
6.	(a) Ability to Appeal	.10	.13*	.29*	.28*	.38*	—						
7.	(b) Appraisal Appealed	-.12	-.15*	-.28*	-.24*	-.33*	-.86*	—					
	Hypothesis 13												
8.	Total Recs Followed	.50*	.50*	.72*	.64*	.76*	.60*	-.56*	—				
	Hypothesis 16a												
9.	Public vs Private Sector	.03	.04	-.11*	-.08	.07	.05	-.06	.06	—			
	Hypothesis 15												
10.	Rating Decrease	.23*	.23*	.44*	.30*	.30*	.17*	-.16*	.41*	-.01	—		
	Hypothesis 17a												
11.	Satisfactory Appraisals	.10	.09	.27*	.01	.21*	.11*	-.07	.27*	.03	.39*	—	
12.	Judicial Decisions	.15*	.15*	.10	.25*	.14*	.01	-.04	.14*	-.15*	-.01	-.21*	—

**p* <.05 (1-tailed).