One Hundred Years of Discrimination Research in JAP:

A Sobering Synopsis

Adrienne Colella, Mikki Hebl, Eden King*

Tulane University, Rice University, George Mason University

Author Note

*Order of Authors is alphabetical reflecting equal contributions on all parts. Correspondence should be sent to Adrienne Colella, A.B. Freeman School of Business, Tulane University, New Orleans, LA, 70118.
Abstract

Employment discrimination—a legal, social, moral, and practical problem—has been a persistent focus of narrow scholarship in the *Journal of Applied Psychology* since its inception. Indeed, this paper identifies the environmental characteristics, conceptual underpinnings, dominant methodologies, research questions and findings across 508 papers published on discrimination in the journal over the last 100 years. Emergent themes document signs of stability and change in three eras: an era wherein discrimination research was itself discriminatory (1917-1969), the heyday of discrimination research (1970-1989), and an era of unsteady progress (1990-2014). This synthesis suggests that, although increasingly sophisticated methodological approaches have been applied to this topic, the targets of focus and theories driving research have largely been static. Additionally, research published on discrimination in the *Journal of Applied Psychology* has often trailed too far behind the times. Specific recommendations for advancing the psychological study of employment discrimination in applied contexts are provided.
One Hundred Years of Discrimination Research in JAP: A Sobering Synopsis

When the first issue of the Journal of Applied Psychology (JAP) was published in 1917:
(1) All three authors of this paper could not vote, (2) It was legal to reject hiring people on the basis of their race, sex, and religion, (3) In some states, it was demanded that Blacks and Whites receive separate schooling, public services, and access to goods and services, and (4) U.S. armed forces were segregated. Yet, paradoxically, the field of industrial and organizational psychology has addressed U.S. employment discrimination for the entirety of the past 100 years\(^1\).

Here we review the body of work on employment discrimination in JAP and thereby offer unique insights that differ from general reviews of (a) the journal that do not focus on discrimination (Casco & Aguinis, 2008), (b) social psychological prejudice research (Duckitt, 1992; 2010), and (c) organizational diversity (Hebl & Avery, 2013). We expand the International Labor Organization (ILO) of the United Nations’ (2015) definition of employment discrimination-- “any distinction, exclusion or preference made on the basis of race, colour, sex, religion, political opinion, national extraction or social origin, which has the effect of nullifying or impairing equality of opportunity or treatment in employment or occupation”-- to explore research on workplace discrimination against any group.

Historical Review: Discussion of Research by Era

To ensure adequate breadth and depth in our review of discrimination research published in the JAP over the past 100 years, we used four systematic strategies. First, we conducted keyword searches using terms such as: discrimination, harassment, bias, group differences, race (ism), sex (ism), ethnic(ity), gender, adverse impact, differential validity, stereotype, prejudice,

\(^1\) Our review is ethnocentric in the sense that it focuses primarily on research and conclusions based on U.S. samples and U.S. definitions of discrimination. This is an artifact of the sample: out of the 508 studies we reviewed, only 17 were clearly conducted on samples outside the U.S. We address this issue in more detail in our conclusions section.
disability, race, obese, gay, and age. Second, we examined the abstracts of every paper that appeared in every issue of *JAP* and identified additional papers that were not captured by the search. Additionally, we reviewed all papers considered to be directly relevant to employment discrimination (see Online Appendix). Third, we mapped impactful *JAP* articles against historical reviews of social and legal landmarks (Cascio & Aguinis, 2008; Salgado, 2001; Salmon et al., 2013) to contextualize the topics of focus and identify gaps or lags in the responsiveness of papers in *JAP* to historical events. Finally, we consider the alignment between the theories and findings specific to employment discrimination research at *JAP* with those that are common in the study of prejudice in other domains of psychology (Duckitt, 1992; 2010) to clarify the relative status of employment discrimination research in *JAP*. We synthesize these approaches to discover key characteristics of relevant eras: social and legal context, conceptual underpinnings, common research questions, dominant methodologies, and general conclusions. In so doing, we cite exemplar studies that spurred new lines of questioning, ended a line of research, or demonstrated methodological innovation. Thus, we present an informative picture of how this journal has contributed to research on employment discrimination (Figures 1-3).

**1917-1969: The Era of Discrimination Research That Was Discriminatory Itself**

With the journal’s inception in 1917 began, too, the publishing of articles that focused on issues related to discrimination (Figure 1). However, such articles contained discrimination not only in their research focus but also, ironically, in the racist and sexist beliefs that guided such research and were evident therein. Egregiously, for instance, Arlitt (1922) tried to “determine the median IQ of all ages of Negroes to show to what extent the inferiority of Negroes is dependent on age.” Later, Garth (1927) reported that, “IQ decreases, roughly speaking, with degree of Indian blood” (p. 273). Attitude scales given to assess how people felt about the treatment of
Black people contained chilling items (e.g., “Negroes do not deserve the privilege of a court trial like other people” p. 374) and responses from White participants (Sumner & Shaed, 1945). Race was not the only target; women, too, were viewed to be inferior to men. Porteus (1924) stated: “I believe that to this superiority of post-pubescent brain growth we can ascribe the greater average general efficiency of males and that herein also lies the fundamental reason why this is and will remain a man-directed and not a woman-governed world—universal suffrage and woman's rights and coeducation notwithstanding. They [women] tend to lack planning capacity, they are irresolute or impulsive, easily confused, excitable, obtrusive, suggestible and simple’ (p. 63).

Yet, in these early studies, despite the prevailing belief that women and Blacks were inferior to men and Whites respectively (e.g., see Duckitt, 1992), evidence began to amass supporting what would be articulated clearly almost 100 years later: such differences are overinflated, and the races and genders are much more similar than different on most physical and psychological variables (e.g., Hyde, 2005). Unfortunately, scientific racism prevailed during this early era and researchers presumed the superiority of White men.

**Social and Legal Context**

In the context of the world wars and the enlistment of women and ethnic minorities in the armed forces, four critically important legal actions emerged during this time period in the U.S. The first was the passage of the 19th amendment, which granted women the right to vote and was the culmination of the first feminist movement and a wave of women who believed that they should be able to have the same basic rights as men. The second legal action was the Immigration Act of 1924, a law that widely restricted the influx of new and diverse people into America until its repeal in 1965 (Stephenson, 1964). The third very significant legal action was the landmark U.S. Supreme Court case, Brown v. Board of Education (1954), in which state-
sponsored segregation in public education was found to be unconstitutional while the country was experiencing the greatest movement in modern American history. Finally, the most important legal actions involving civil rights occurred late in this era, more than a decade after Americans witnessed the atrocities of the Holocaust. The Civil Rights Act of 1957, which created an independent U.S. Commission on Civil Rights, lead the way to the landmark Civil Rights Act (CRA) of 1964. Though there was not a great deal of power in enforcing this legislation initially, it emerged as the single most meaningful piece of legislation to date forbidding discrimination based on race, color, religion, sex, and national origin.

**Conceptual Underpinnings**

The conceptual forces driving this period of research involved race theories. As revealed in Haller (1971), “almost the whole of scientific thought in both America and Europe…accepted race inferiority” (p. 77). It was that this belief manifested into a theoretical approach guiding much of the research. Articles in *JAP* revealed the predictive ability that race, gender, and other variables ostensibly had on intelligence and other outcomes. This is reflected by Porteus (1924), “…the Japanese are considerably superior to the Chinese and Portuguese, and males of every race are superior to the females” (p. 71). Wildly overlooked during this timeframe was the role that the situation or environment could play, such as fact that White individuals also received better schooling, had higher levels of SES, and had far more opportunities than did minority individuals. Duckitt et al. (1992) lists other common conceptual foci during this time period and particularly cites viewing prejudice as a problem, identifying universal processes underlying prejudice, and identifying the prejudice-prone personality. However, an investigation of the articles published in *JAP* does not reveal that these issues were actually taken up by *JAP* authors.

**Questions Asked in JAP Articles**
Of the 139 papers we identified during this time period, papers targeted race (47%), gender (43%), age (<8%), SES (<8%), intellectual disability (<6%), and characteristics such as size, physical disability, or marital status (10%). A total of 22% of the studies reported on more than one social identity group, which was most commonly race and gender. The research questions being asked in *JAP* articles during this timeframe reflect a belief about race and gender inferiority. Such beliefs are reflected in the titles of articles published in *JAP* that would be unthinkable today: “The moron as a war problem (Mateer, 1917),” “White, Indian, and Negro work curves” (Garth, 1921), “The intelligence of mixed-blood Indians” (Garth, 1927), “The mentality of the Chinese and Japanese” (Hsiao, 1929), “Patterns of Jewish temperament” (Sward, 1935), “Idiot, imbecile, and moron” (Doll, 1936), “Menstruation and industrial efficiency” (Smith, 1950), and “An easier ‘male’ mechanical test for use with women” (Mollenkopf, 1957). The majority of such studies, particularly those conducted in the first half of the 20th century, did not question whether such differences existed. Rather, researchers attempted to quantify the amount, extent, and nature of such differences between groups of people, with the White male serving as the comparison and presumed superior group.

**Dominant Methodologies**

During the timeframe of 1919-1925, a widely regarded mathematical statistician, Sir Ronald Fisher, pioneered many of the principles regarding the study of the analysis of variance (see Conniffe, 1990). Nevertheless, the majority of research published in *JAP* during this time was far less advanced. One predominant methodology, for example, involved the use of descriptive statistics. Examples of this include measuring the “intelligence of the full-blood Indian” (Garth, Smith, & Abell, 1928), “the mental growth of dull Italian children” (Wheeler, 1932), or the “group manual dexterity in women” (Comrey & Deskin, 1954).
A second predominant methodology was quantifying differences between groups of individuals; hence, it was common to identify two groups (e.g., Blacks and Whites: Eagleson, 1938; men and women, different native American groups from Alaska: Eells, 1933: Chinese and Japanese: Hsaiao, 1929) and quantify the differences that emerged when they took some intelligence, aptitude, personality, or other test (e.g., handwriting, cranial capacity).

A third and related methodology during this time involved either correlating the scores of particular groups on multiple scales, or comparing their overall scores to a dominant comparison group. Examples of the former type of methodology include studies looking at the “the personality of Indians” (Garth & Garth, 1937), “Jewish temperament “ (Sward, 1935); and examining “vocational and personality tests with deaf” (Lyon, 1934). Studies of the latter type include “intercorrelations among metabolic rate, vital capacity, blood pressure, intelligence, scholarship, personality and other measures on university women” (Patrick & Rowles, 1933), how “worker characteristics relate to age” (Bowers, 1952), and “the relation of scholastic aptitude to socioeconomic status” (Washburne & Andrew, 1954).

A fourth less common, but still prevalent, methodology was to explore intergroup attitudes. In Davenport and Craytor’s (1923) study examining the “comparative social traits of various races,” teachers rated students of six different ethnicities (e.g., Russian, Irish) on a number of different traits; German students were rated highest in leadership and sense of humor ratings and Italian students were rated lowest. In other attitudinal research, Link (1944) reported results from a nation-wide survey that assessed the prejudicial attitudes individuals held toward various groups of people. Moreover, Tuckman and Lorge (1952) reported on the attitudes that people hold toward older workers.
In sum, this timeframe saw a reliance on measurement, comparative, and descriptive studies as the principal methodologies. One unique study by Athey, Coleman, and Tang (1960), involved two quasi-experimental studies in the field. The researchers found that both peers and homeowners toned down their racial discrimination when they were asked about it by Black or Asian (versus White) interviewers. The fact that this experimental field methodology study was conducted in 1960 is profound but an exception to the rule.

**Themes of Published Findings**

Several themes emerge from studies published in *JAP* during this timeframe. First, the prevailing belief was that certain races and peoples were superior to others, particularly on measures of intelligence and other skills and abilities. Researchers spent enormous efforts researching and *JAP* devoted significant pages to results that supported such findings, whether wholly or partially. As stated by the opening sentence of Garth (1921), “the investigator working in the field of racial psychology is looking either for similarities or differences, or perhaps for both.” (p. 14). Moreover, even when null results were found, the absence of such differences was reported. Across more than 50 years, quite a few null differences were found, thereby providing the foundation to scrutinize the general ideology. Second, a large number of studies compared men and women on a variety of different dimensions or studied women only. From these studies, we learn that the genders tend to have similar food aversions but that “social pressures exist which force males to repeat experience with disliked food but which permit females to retain habits of rejection” (Wallen, 1943); Chinese girls score higher on Binet tests than do Chinese boys (Yeung, 1921), penmanship quality is similar between the genders (Tenwolde, 1934), that men like “shady” jokes better than do women (Omwake, 1937). These studies provide evidence of early reinforcements of gender roles, maintaining gender-related biases, stereotypes, and
discrimination. Finally, one important current that emerged during this time was assessing and exploring how prejudice might be compounded or alleviated. For instance, a nation-wide survey by Link (1944) showed that the majority of people believe that churches and Sunday schools could do the most good in reducing prejudice and hatred toward people of lower classes (p. 367).

Overall, the research published during this era largely confirms that a paradigm shift was needed for JAP. The dawning of the Civil Rights Act, the lessons learned from World War II and the lack of guidance for gender neutrality and ethical treatment of human beings in scientific research are about to collide, and bring such a paradigm shift forth.


While research in the 1960s was slow to address the ramifications of Title VII of the CRA, studies related to employment discrimination flooded into the literature during the 1970s and 1980s (n = 187). It seems that while great strides were made in anti-discrimination legislation and the civil rights movement during the 1960’s, it was not until the 1970s that enforcement of legislation came to be a major issue to employers. It was then that questions arose that were of interest to the field of applied psychology, spurring a great deal of research.

Social and Legal Context

The 1970s were an era of upheaval in light of the war in Vietnam, continuation of the civil rights movement and the feminist movement, and the vast entry of women into the workforce in the U.S. Additionally in 1972, Congress amended the CRA to make it more effective in preventing employment discrimination by passing the Equal Employment Opportunity Act (EEOA) (EEOC, 2015). The ability of the EEOC to enforce Title VII was further strengthened by the adoption of Uniform Guidelines on Employee Selection Procedures (UGESP, 1978). The Rehabilitation Act (1973) and the Pregnancy Discrimination Act (1978)
were passed to protect federal employees from discrimination based on disability and pregnancy. Sixteen major Supreme Court rulings increased the scope of Title VII. For example, in *Griggs v. Duke Power Company* (1971) the court ruled that when employers use a neutral (not blatantly discriminatory) test or policy in selection that adversely affects women or minorities, then the employer is responsible for demonstrating the business necessity of using the test or policy.

The 1980s are usually depicted as optimistic, prosperous, and conservative (Collins, 2007). However, the issue of rising income inequality and the rising number of homeless in the U.S. (Collins, 2007) indicated that disparities were growing. The AIDS epidemic also appeared, which brought about another wave of activism in the gay community. The Age Discrimination in Employment Amendments (1986) removed the 70-year old cap on protection under the Act and the Immigration Reform and Control Act (1986) set forth that employers can be sanctioned for hiring undocumented immigrants. In 1980, the EEOC issued Interim Guidelines on Sexual Harassment stating that sex-related intimidation on the job or creating a sexually charged hostile work environment is unlawful under Title VII. Yet, the Supreme Court narrowed Title VII protections and enforcement was lightened (see Eskridge, 1991). For example, in *Wards Cove Packing Co. v. Antonio* (1989), the Court ruled that the employee must show not only disparate impact but also shifted the burden of proof from the employer to the plaintiff. Also, sexual harassment was ruled to be in violation of Title VII in *Meritor Savings Bank v. Vinson* (1986). *JAP* authors were very timely in addressing some of these trends and late to the game on others.

**Conceptual Underpinnings**

The prevailing conceptual explanations of prejudice and discrimination during the 1970s focused on intergroup conflicts of interests, structural power differentials, and institutionalized racism (Duckitt, 2010). A sociological perspective, rather than a psychological perspective, was
adopted in the social sciences. The 1980s saw a shift in explanations of discrimination to the cognitive perspective and a focus on social categorization and identity (Duckitt, 2010). These perspectives were more influential on shaping research done in the JAP. Stereotyping, a cognitive construct, was used throughout this time period to explain differential treatment. Prominent theories influencing research during this era were Byrne’s (1971) Similarity Attraction Paradigm, McConahay and Hough’s (1976) Symbolic Racism theory, Tajfel and Turner’s Theory of Inter-Group Conflict, Kanter’s (1977) work on tokenism, Heilman’s (1983) stereotypical fit model, McConahay’s (1986) Modern Racism theory; Tajfel and Turner’s (1986) Social Identity theory, Eagley’s (1987) Role Congruity theory, and Gaertner and Dovidio’s (1986) Aversive Racism theory. Overall, most discrimination research in this time period relied on stereotypes and role conflict as explanatory constructs for discrimination (c.f., Terborg, 1976). Other research on test fairness focused on psychometric theory and statistical reasoning. A few studies examining relational demography effects (a term coined in the late 1980s; e.g., Turban & Jones, 1989) relied on Byrne’s (1971) similarity attraction paradigm.

Questions Asked in JAP Articles

Most studies during this period focused on sex or gender (59%) and race/ethnicity (35%). Almost all studies looking at race considered only White versus Black targets (28% of all studies). Another frequently examined issue (14.4%) examined how women were evaluated or perceived as leaders (e.g. Lord, Phillips, & Rush, 1980), stereotypes about women as managers (e.g., Schein, 1973), and whether women are as likely as men to emerge as leaders (e.g., Carbonell, 1984). To further attest to the popularity of this topic, the top three “most cited” articles in JAP from this period are longitudinal studies examining the overlap between sex role stereotypes and stereotypes of managers (Figure 2).
Research concerning test fairness also emerged (13%) during this period and focused on differential validity, differential prediction, and single group validity (e.g., Hunter & Schmidt, 1978). Indeed, there was an eight article debate running through 1977-78 volumes of *JAP* over the presence of differential prediction/single group validity and arguments about how one defines and assesses them (e.g., Hunter & Schmidt, 1978, Schmitt, Coyle, & Mellon, 1978). Clearly this was a hot topic in the late 1970s. A final issue that received attention (13%) during this period was the difference in job-related attitudes or values of various groups (usually men vs. women and Whites vs. Blacks) that may account for their different experiences. On the surface this may not seem like discrimination research, but these studies pointed to unfair treatment or discrimination as the basis for group differences.

Two very practical issues also were given attention in the pages of *JAP* during this period. Nine studies examined how adverse impact could be reduced during selection (e.g., Ironson, Guion, & Ostrander, 1982). Seven studies addressed the issue of comparable worth by examining bias in job evaluation methods (e.g., Doverspike, Carlisi, Barrett, & Alexander, 1983). Beyond these studies, it is surprising that so few additional ones (N = 8) examined bias in treatment and work experiences of people from different groups. It is also surprising that, given that affirmative action was a big issue during the 1970s, only seven studies addressed this. Despite legal and social events during this time period, there was no research on LGBT issues, sexual or racial harassment, strategies to reduce discrimination beyond testing and adverse impact, or work exploring compensation discrimination.

**Dominant Methodologies**

This era involved the laboratory experiment (26% of total studies) and “paper people” studies (Murphy, Herr, Lockhart & Eammon, 1986). The general paradigm was that participants
received information about bogus employees or job candidates with varying demographic characteristics holding performance constant. This basic paradigm was followed in performance evaluation studies (e.g., Dipboye, Fromkin & Wiback, 1975), interview studies (e.g., Raza & Carpenter, 1987), and selection studies (e.g., Stone & Stone, 1987). A less common form of the lab experiment was to actually have participants engage in a work simulation or some other type of actual interaction with the target person (e.g., Richards & Jaffee, 1972).

The generalizability of the findings of this type of experimental paradigm to real world evaluation situations was criticized immediately (e.g., Dipboye, 1985; Wendelkin & Inn, 1981) based on its lack of ecological fidelity. However, a meta-analysis of performance rating studies (N=111) found that the average effect size for demographic effects to be $d = .25$ in paper people studies and $d = .19$ in observational studies (Murphy et al., 1986). In addition, a meta-analysis of race effects on performance evaluations (N=81 studies) found larger bias effects against Black ratees in field studies ($d = .39$) compared to laboratory studies ($d = .07$) (Kraiger & Ford, 1985). Thus, at this time it appeared that lab experiments did not inflate bias effects, a finding that is continuously challenged (Landy, 2008).

Another methodology employed during this time was field studies comparing ratings, behaviors, traits, values or attitudes across different groups of employees. Fourteen studies made sex or race comparisons of performance ratings (e.g., Thompson & Thompson, 1985), interview evaluations (e.g., Parsons & Liden, 1984), or some other type of selection evaluation (e.g., Abrahms & Atwater, 1977), with the best of these studies controlling for some measure of objective performance (e.g., Pulakos et al., 1989). Seven studies assessed experiences and perceptions of the workplace (e.g., Gordon & Hall, 1974), five assessed traits (e.g., Brenner, 1982), and three studies each assessed work-related interests or behavior (e.g., Izraeli, 1985).
A total of 6.4% of studies in JAP were psychometric studies (10.2%), which were either conceptual (Boehm, 1977), re-analyzed old data with new assumptions (e.g., Hunter & Schmidt, 1978), or simulation studies (e.g., Pritchard, Maxwell, & Jordon, 1984). Almost all these studies were related to disparate prediction and compared validities and other aspects of prediction among different groups. Finally, there were only five purely conceptual papers (e.g., Linn, 1977) and reviews (Terborg, 1977) which is perhaps why, up to this point, the papers in JAP made few theoretical developments in the field of employment discrimination.

Themes of Published Findings

This was an era of mixed findings perhaps due, in part, to the simplicity of the questions asked. Few moderators of bias effects were examined systematically, and rarely were explanatory mediator effects studied. Studies on gender and racial biases in performance evaluations revealed mixed findings: some found that women/Blacks were rated more negatively than men/Whites (gender: Bartol & Butterfield, 1976; race: Pulakos et al., 1989), some found no differences (e.g., gender: Rosen & Jerdee, 1973; race: Turban & Jones, 1988), and others found that women/Blacks were rated more positively than men/Whites (gender; Bigoness, 1976; race: Wendelkin & Inn, 1981). For gender, the only moderator examined with any consistency was the gender of the rater but the evidence was not strong (e.g., Schmitt & Hill, 1977; Wexley & Pulakos, 1982). For race, rater X ratee race interactions with Whites favoring Whites and Blacks favoring Blacks were consistent (e.g., Bigoness, 1976; Turban & Jones, 1989). Kraiger and Ford (1985)’s meta-analysis (N=81) supported the finding of rater X ratee effects such that the average effect size for White raters was \( d = .37 \) (in favor of Whites) and for Black raters \( d = -.45 \) (in favor of Black rates). Some studies also looked at age bias in evaluations (e.g., Raza & Carpenter, 1987; Rosen & Jerdee, 1976; Turban & Jones, 1989) with mixed results. During this
era, two meta-analyses (McEvoy & Cascio, 1989; Waldman & Avolio, 1986) published in *JAP* assessed potential age bias in performance evaluations, the more comprehensive of which (N=96 studies), McEvoy & Cascio (1989) found no relationship between age and productivity (r = .07, ns) or objective ratings (r = .03, ns), indicating that bias in performance evaluations is not an issue and refuting the assumption that performance declines with age.

Several themes emerged from the 14.4% of discrimination articles addressing the issue of women as managers. First, male, but not female, stereotypes coincide with manager stereotypes (Heilman, Block, Martell, & Simon, Schein, 1973). Second, women might not perceive themselves as being “management material” since they also hold stereotypes (Schein, 1973). Third, there was little evidence of gender differences in effectiveness in managerial roles. Considering the main effect of sex on ratings of supervisors, null results were the most common (e.g., Peters et al., 1984). But leadership styles did make a difference; for instance, female leaders are rewarded more for showing people-oriented, nurturing, or low power leadership styles and punished for using more aggressive, autocratic, agentic styles than men (e.g., Bartol & Butterfield, 1976; Izraeli, 1985).

Test bias was also a hot topic in the 1970s and this work centered on the notion of differential prediction and single group validity. An eight-article debate on test-bias appeared in *JAP* in the mid-70s, which Linn (1978) summarized as covering several issues: the existence of single group validity, definitions of differential prediction, how data from previous studies should be analyzed (e.g., the independence of data); the relevance of single group validity, practical differences in validity, and the need to quit focusing on validities and focus on the entire prediction system (e.g., intercepts, etc.). The general conclusion was that there was little evidence of differential prediction and when it did occur, there was over-prediction for minority
groups. The notion of single group validity was dropped. Subsequent research also focused on problems with assessing differential prediction such as small sample sizes (Trattner & O’Leary, 1980) and measurement equivalence among groups (Drasgow & Kang, 1984).

Finally, many studies examined group differences in a wide variety of attitudes. Themes across these diverse attitudes are unclear, with the exception of consistent findings suggesting that Blacks were less satisfied with their jobs than Whites (e.g., O’Reilly & Roberts, 1973). Konar (1981) tested potential explanations for this discrepancy and found that racial differences in job satisfaction could largely be explained by differences in work experiences. Another theme was that there was little difference between White and Black employees on the importance of higher-order, intrinsic job characteristics, but Blacks often placed more importance on extrinsic factors such as pay that could be due to large race gaps in these experiences (Moch, 1980).

In sum, discrimination research in JAP during this period came into its heyday, experiencing the paradigm shift called for at the end of the 1960s. Research identified the bias in evaluations, women in leadership roles and the discrimination they may face, test fairness, differences in job attitudes due to racial differences in work treatment, and the assumption that work meant something different to women than it did to men.

1990-2014: The Era of Unsteady Progress

The social upheaval of the 1960s and 70s has evolved into smaller, subtler shifts forward and backward in the social context and in the applied psychological research findings that emerge in lagged response. Indeed, the themes emerging from our review of discrimination research in JAP between 1990 and 2014 imply unsteady progress (Figure 3).

Social and Legal Context
This time frame was characterized by both remarkable social challenges and social progress in the U.S. The Gulf War began in 1991 and was followed by the terrorist attacks of September 11th, 2001 and incursions into Afghanistan and Iraq. President Clinton signed the “don’t ask, don’t tell” policy in 1993, which prohibited openly gay men and lesbians from serving in the military. The political sphere itself was also an important avenue through which social change emerged in this era. For example, dialogue regarding sexual harassment became predominant in the early ‘90s when the judiciary committee interviewed Supreme Court candidate (and later Justice) Clarence Thomas and his employee Anita Hill about her allegations of sexual harassment. It was also during this time (2007) that the first female Speaker of the House—Nancy Pelosi—was elected. Perhaps most significantly, Barak Obama was elected President of the United States in 2003. Yet, this era also was bookended by strikingly similar reactions to racially charged incidents involving police. In 1992, riots erupted when police officers were acquitted in the beating of Rodney King. Similarly, in 2014, protests emerged when a grand jury did not charge the officer involved in shooting death of Michael Brown.

Specific to the employment context, five particularly important legal actions emerged during this era. First, in 1990, the Americans with Disabilities Act was passed, requiring employers to provide reasonable accommodations that are not significantly difficult or expensive to employees with disabilities. Second, the Civil Rights Act of 1991 allowed jury trials and the recovery of compensatory and punitive damages for intentional discrimination, representing the evolution of civil rights claims on the basis of unintentional discrimination. For example, selection practices that result in adverse impact can now be challenged regardless of their intent. Third, the Federal Medical Leave Act was passed in 1993, guaranteeing that workers who have been employed in a full-time, salaried position by a company with more than 50 employees for a
year must be given up to 12 weeks of unpaid leave. Fourth, the cases of Grutter v. Bollinger and Gratz v. Bollinger (2003) considered the use of race in university admissions and concluded that race-conscious admissions procedures can be consistent with a compelling interest to promote diversity, clarifying that race can be considered in “narrowly tailored” decisions but quotas are unconstitutional. Interestingly, the decision of the court specified that a race-neutral approach should be implemented when it is no longer necessary to ensure equal access. Fifth, President Obama’s first formal action in office was to sign the Lily Ledbetter Fair Pay Act to amend the Civil Rights Act to specify that the statute of limitations of compensation discrimination claims restarts with each paycheck. It was in the context of these social and legal steps forward and backward that JAP authors used a variety of theoretical bases to explore relevant questions.

**Conceptual Underpinnings**

The underlying theoretical basis for much of the research conducted during this time period is based in social identity and social categorization theories (Tajfel & Turner, 1979). That is, papers published in JAP on employment discrimination often refer to intergroup biases that emerge as a result of social group memberships and their importance in defining our sense of selves and others. These categories trigger the formation of stereotypes, the development of prejudice, and the enactment of discrimination. Most of the papers in this era present refinements or extensions of this overarching theoretical framework. For example, the sizable work by Heilman (e.g., Heilman & Okiomoto, 2007; 2008) relies on a lack of fit perspective wherein the stereotypes associated with women are inconsistent with the expectations of good leaders or good workers in masculine positions. In addition, studies that examine the role of social identity characteristics in selection contexts discuss not only stereotypes but also the role of similarity-attraction processes (Lin, 1992). As another example, work on attitudes towards Affirmative
Action (e.g., Harrison et al., 2006) relies on the motivational perspective of self-interest. Work on outcomes of sexual harassment and discrimination often relies on models of stress and coping to define these negative workplace experiences as stressors (e.g., Fitzgerald et al., 1997). These perspectives have yielded substantial knowledge and understanding, but we cannot know what new discoveries might be enabled with a wider variety of viewpoints.

In addition, not captured by the aforementioned theories are the two basic explanations for adverse impact and differential validity. On the one hand, it is argued that genuine differences in the phenomena of interest (e.g., intelligence, physical ability) account for group differences in performance on selection tests and/or performance itself. This is consistent with the findings of, for example, Arvey and colleagues (1992), who found that male and female police offers demonstrated different levels of physical abilities on both tests and performance. On the other hand, it is also argued that little or no genuine differences exist beyond those that are created as a result of biases (e.g., by biased testing procedures, stereotype threat experiences, bias in performance appraisal). For example, Chan and Schmitt (1997) found that video-based situational judgment tests yielded smaller Black-White differences than traditional paper-and-pencil tests. These conceptual—and to some extent, philosophical—perspectives are often the implicit focus of papers published on this topic.

**Questions Asked in JAP Articles**

The primary targets of focus in research during this time frame was gender (39% of papers) or race/ethnicity (27% of papers), with 12.6% reporting on more than one social identity group and very few studies (less than 5%) published on age, disability, sexual orientation, pregnancy, obesity, and other characteristics. Yet, the introduction of these other social identity categories is in and of itself notable; it was not until this era that these targets were included in
On the positive side, the papers of Button (2001), Ragins and Cornwell (2001), and Griffith and Hebl (2002) represent the first studies on sexual orientation published in the journal. These papers appeared after “don’t ask, don’t tell”, but before sweeping changes in same-sex marriage equality and thus represent an important example of timely and relevant scholarship in the journal. On the negative side, however, the small number of papers that consider disabilities (n = 3) during this era seems at odds with the passage of the Americans with Disabilities Act in 1990.

The overwhelming majority of the papers reviewed in this era (72%) pertain to differential treatment and experiences across social identity groups, while the remaining (28%) consider differential validity and adverse impact. The most commonly studied topics on differential experiences were: harassment (n = 19), affirmative action (n = 16), and bias in selection contexts (e.g., interviews; n = 15), performance ratings (n = 12), career progression (n = 11), and interpersonal treatment (n = 10). Several papers also covered the topics of: legal issues (n = 5), leadership (n = 4), compensation (n = 4), identity management (n = 4), mentoring (n = 4), and self-fulfilling prophecies (3). It is interesting that the most frequent topics of papers were not necessarily those that have received the highest citations; for instance, several of the most impactful papers published in this era concern gender and mentoring (Figure 3).

Harassment, the most frequent paper topic, was spurred by the impactful work of Fitzgerald, Drasgow, Hulin, Gelfand, and Magley (1997), which represented the first effort to systematically identify antecedents and consequences of sexual harassment. A continuous stream of subsequent works through 2007 (e.g., Berdhal, 2007; Schneider, Swan, & Fitzgerald, 1998; Sims, Drasgow, & Fitzgerald, 2005) explored these factors and boundary conditions (e.g., legal constraints, coping) that reduce sexual harassment and its consequences. A subset of these papers
examined racial forms of harassment (e.g., Berdahl & Moore, 2006; Blanchard, Crandall, Brigham, & Vaughn, 1994; Raver & Nishii, 2010).

Affirmative action as a research topic also was popular during this time period. Two basic questions appeared in these studies: (1) What are people’s attitudes toward different types of affirmative action programs? (e.g., Harrison, Kravitz, Mayer, Leslie, & Lev-Arey, 2006, and (2) What are the implications for attitudes toward and self-perceptions of beneficiaries of these programs (e.g., Heilman, 1991; Slaughter, Sinar, & Bachioci, 2002)?

The questions of interest in the context of bias in selection (most of which were published before 2000) focused on the roles of demographic similarity and stereotypes on employment decisions. The targets of focus within this general topic were quite diverse (age: Finkelstein et al., 1995, race and gender: Macrae, 1994; attractiveness: Marlowe, Schneider, & Nelson, 1996; obesity: Pingitore et al., 1994). The same questions were asked in reference to performance assessment during the first half of this era, wherein papers addressed the role of demographic similarity and stereotypes in performance ratings (disability: Colella, DeNisi, & Varma, 1998; gender: Sackett et al., 1991; race: Roth, Huffcut, & Bobko, 2003). Yet, in the second half of this era, the questions about discrimination in performance ratings were a bit more nuanced. For example, many studies by Heilman and colleagues (e.g., Heilman & Okimoto, 2007; 2008) considered the ways in which gendered positions or tasks can exacerbate and attenuate the extent of gender bias in performance ratings.

Based largely in concerns about data suggesting that women are stopped by a “glass ceiling” in their advancement, a number of papers explored explanations for gender differences in career progression (e.g., Lyness & Thompson, 1997; Stroh, Brett, & Reilly, 1992). These
papers compare the experiences of men and women across hierarchical levels to explain why women are underrepresented at the top of the corporate ladder.

Particularly in the latter half of this era, scholars shifted from a focus on discrimination in formal employment decisions (such as selection, performance, promotions) toward a focus on discrimination that occurs in interpersonal interactions. That is, rather than asking how stereotypes influence decisions, these studies (e.g., Ragins & Cornwell, 2001; Hebl et al., 2007; Lim & Cortina, 2005) ask about the ways stereotypes influence interpersonal behaviors encountered by targets of prejudice at work.

The set of studies that explore differential validity and adverse impact raise questions about three general topics. First, these papers generate questions about the existence of genuine group differences in abilities and characteristics and in performance (e.g., Arvey et al., 1992; DeShon et al., 1998; Roth, Buster, & Bobko, 2011). Second, studies contrast adverse impact resulting from a variety of methods and tools (e.g., Chan & Schmitt, 1997). Third, another set of papers offer or assess predictive models that balance concerns about adverse impact with a desire to maximize validity (e.g., Decorte et al., 2006; Finch, Edwards, & Wallace, 2009). Each of these questions was addressed from a variety of methodological approaches.

**Dominant Methodologies**

This era of employment discrimination research was marked by considerable diversity in the methodological approaches taken in *JAP*. Nearly half of these papers (49%) presented data from the field, ranging from cross-sectional and longitudinal employee surveys to archival indicators of career progression to applicant demographics and test scores and performance appraisals. About a quarter (24.2%) of papers were based on laboratory experiments that manipulated social identity characteristics (such as race or gender) and assessed the causal
implications of these characteristics on selection decisions, performance ratings, or promotion recommendations. In addition, a meaningful number of meta-analyses (7.5%) were presented as evidence regarding employment discrimination (particularly as it pertains to adverse impact in selection contexts) has accumulated. About 5% of papers in this era included multiple methods (e.g., both survey and experimental data), and others used simulations and Monte Carlo designs.

It is remarkable that the new century of *JAP* research has included a number of genuine field experiments (4.3%). These studies generally manipulate the social identity characteristics of job seekers who apply for actual jobs in face-to-face interactions (e.g., gay and lesbian applicants: Singletary & Hebl, 2009; obese people: Agerstrom et al.; 2013; pregnant applicants; Botsford Morgan et al., 2014). The introduction of this paradigm is significant because it can overcome methodological criticisms of survey and laboratory research and because it captures the contemporary, subtle forms of discrimination that emerge in interpersonal contexts.

**Themes of Published Findings**

The themes that emerged from studies in this era can be summarized with two general conclusions. First, these studies demonstrate that employment discrimination persists toward a number of disadvantaged targets and has critical consequences. Second, attempts to resolve such biases themselves can result in mixed outcomes. Indeed, the studies published in *JAP* in this era of unsteady progress document ongoing experiences of sexual and racial harassment as well as more general forms of discrimination in employment decisions and interpersonal interactions. These findings are consistent with an earlier trend in the social psychological literature that places prejudice as an inevitable outcome of social categorization (Duckitt, 1992). Moreover, the findings during this era show that experiences of harassment and discrimination can be detrimental to targets’ mental and physical health, job attitudes, performance, and career
progress. What is less clear in the papers we reviewed is the extent to which such experiences ultimately create or underlie adverse impact or differential validity.

Additionally, research during this time show that efforts to reduce adverse impact and employment discrimination seem to be, at best, moderately effective and, at worst, destructive. Work on attitudes toward Affirmative Action shows that people inaccurately believe that its beneficiaries are undeserving of the opportunities that they receive (Harrison et al; 2006; Heilman, Block, & Lucas, 1992). This stigma translates into beneficiaries’ own perceptions of their insufficient competence (Heilman et al., 1991). Exposure to others’ condemning racism can reduce discrimination, but hearing others’ condone racism has the opposite, bolstering effect (Blanchard et al., 1994). Mentoring programs can help women and minority employees achieve career outcomes comparable to their colleagues (Dreher, 1996), but really mostly when they are paired with White male mentors. Women can overcome stereotypes of incompetence but end up being disliked (Heilman & Okimoto, 2008). Only men benefit in terms of compensation from external market strategies (Brett & Stroh, 1997). There is also mixed evidence regarding strategies to reduce adverse impact; some studies show that alternative tests or methods reduce adverse outcomes (e.g., Chan & Schmitt, 1997; Huffcut & Roth, 1997), while other studies show little difference across approaches (e.g., Ryan et al., 1998). Overall, the research published in this era confirms the problem of employment discrimination but fails to provide clear direction for its resolution. The complexity of the cognitive and motivational processes underlying bias, and the continuing social, organizational, and political structures, perpetuate such challenges.

Reflections: The Past 100 Years and the Next 100 Years

The journey in conducting this review of almost 100 years of research on discrimination was sobering, fascinating, and hopeful. Based on our review, several general insights can be
gleaned: (1) discriminatory treatment exists against women and racial/ethnic minorities but is
moderated by complex conditions; 2) adverse impact results from tests (particularly cognitive
ability tests), we are still debating about differential prediction, our efforts to ameliorate this have
been mixed, and efforts to understand why it occurs have been almost non-existent; and 3) and
we have learned little about how organizations can combat employment discrimination. In
looking toward the future, these reflective insights necessitate the following avenues for future
work.

**Wider Range of Discriminated Employees**

One of the most striking aspects of our review was to see that race and gender were
present even at the inception of *JAP*. Interest in this topic revealed that prior to 1960, 3.7% of
*JAP* articles addressed issues related to discrimination, compared to 8.6% between 1960 and
1989, and 4.7% from 1989 until 2015. We encourage more research on the Black and female
experience in the workplace but also hope researchers will extend their focus beyond these
important groups. In particular, there has been a glaring absence from a great deal of the *JAP*
literature is a focus on discrimination more globally. Less than 4% of the discrimination studies
in *JAP* were conducted outside the U.S. The types of discrimination that people experience in the
workforce may be dramatically different than that experienced by international employees (e.g.,
child labor, regulations against men and women working together, increased rates of sexual
harassment). Indeed, discrimination is defined differently both legally and culturally across the
globe. Heightened attention on more global workplace concerns regarding discrimination, at
least up to this point, have been beyond the purview of *JAP*. Clearly the readers of *JAP* represent
a diverse international community; our research must better align with its global audience.
Discrimination research in the next century must broaden the focus not only on experiences outside the U.S. but also on a wider swath of those within the U.S. workforce. Too little attention has been paid to the experiences of employees who are Hispanic/Latino (and who will display a 75% increase in the workforce between 2020 and 2034; IHS, Inc., 2015), Asians, employees who are LGBT, employees from various religious backgrounds, those who are low in SES, veterans, and those with physical and/or mental disabilities). We hope that future issues of *JAP* include these understudied groups.

**Varying Methodology**

Our review shows that most studies were not rigorous until the 1970’s; rather, they were descriptive and/or used simple comparison techniques without sophisticated statistical techniques. From a methodological viewpoint, discrimination has not been an easy topic to study; challenges include getting real world samples, identifying people as members of an identity group, and social desirability effects. The future holds particular promise if researchers unveil new approaches to study discrimination including “big data” opportunities for detecting biases, longitudinal research, and multi-level methodologies and analysis techniques. Employment discrimination is certainly a result of multi-level factors and occurs over time, yet research in *JAP* has not been multi-level or longitudinal. Such approaches allow for consideration of how discrimination processes may change over time and how discrimination can be conceptualized and analyzed at the individual, group, and organizational level. Future research also will benefit from relying upon a range of methodologies, multi-level methodologies, and triangulation of findings (Leslie, King, Bradley, & Hebl, 2008).

**More Theory**
Although the methodology improved over the last century, there generally has been too little theory-advancing work published in *JAP*. Most of the theories that appeared in the journal come from other fields. Part of this may simply be the focus of the journal, which is inherently applied. Whatever the reason, theory has been lacking with some very few notable exceptions such as the sexual harassment model of Fitzgerald et al. (1997). This likely exacerbates a continuing problem of the “black box issue.” That is, employment discrimination researchers over the course of the last century have not done a great amount of work examining the cognitive and emotional processes that precede discrimination. Progress over the next several decades would entail examining this “black box” and focusing on why workplace discrimination occurs.

*More Cutting Edge Issues*

Our review reveals that *JAP* has been behind the times on most discrimination issues. At present, for instance, veterans are seeking employment at unprecedented rates; yet, not a single publication has focused on the potential discrimination they face. The rise in Hispanics in the workforce is not complemented by a rise in research on their workplace experiences. We hope researchers and reviewers respond to the overly long lag time that happens between events and research that responds to these. For example, research is needed to understand the implications of the 2014 implementation of Section 503 of the Rehabilitation Act (1973) which requires government contractors to have a 7% hiring goal for persons with disabilities and to ask people to self-identify as having a disability pre-hire. As another example, police shootings of unarmed Black men necessitate the understanding of bias in policing. We are hopeful that researchers will answer this call to engage more quickly in attention to such critical issues.

*Organizational Level Strategies to Reduce Discrimination*
In recent years, researchers have begun to illuminate some effective individual-strategies to reduce workplace discrimination. For instance, we know that acknowledging a stigma, increasing one’s positivity and likability, and de-individuating oneself from stereotypical associations with a stigmatized status are at least somewhat effective strategies for targets to pursue. However, there is clearly much more work to do to disentangle why, when, and how these strategies are effective, and we hope researchers will continue to clarify these critical details and particularly pay attention to organizational level strategies. For instance, research on empirically tested, effective diversity training is in great need and entirely absent from JAP.

**Conclusion**

In the past 100 years, scholars writing in JAP have come a long way in addressing employment discrimination. Early researchers attempted (largely unsuccessfully) to identify reasons that demographic differences should be used to discriminate. Later, we documented how people were being evaluated and focused attention on adverse impact. Recently, we have distinguished between different types of discrimination (e.g., overt versus subtle) and have begun to shown that bias can be tempered. These are big strides but we have a long way to go in expanding our understanding beyond sex and race, asking different questions, employing different methodologies and multi-level considerations, and measuring discrimination beyond U.S. boundaries. We further hope that JAP’s future will be one in which discrimination research is done with greater attention to pressing current events and not so delayed in responding to changing mores, laws, and social events. We remain optimistic about the future of discrimination research in JAP; indeed, equity in the workplace depends upon it.
References


http://dx.doi.org/10.1037/a0013775


http://dx.doi.org/10.1037/0021-9010.82.4.578


doi: 10.1037/h0036447


Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?:
the implied communality deficit. *Journal of Applied Psychology, 92*(1), 81-92. doi:
http://dx.doi.org/10.1037/0021-9010.92.1.81

employment decisions. *Journal of Applied Psychology, 93*(1), 189-198. doi:
http://dx.doi.org/10.1037/0021-9010.93.1.189

Current characterizations of men, women, and managers. *Journal of Applied Psychology,
74*(6), 935-942. doi: 10.1037/0021-9010.74.6.935

13*(1), 9-31. doi: 10.1037/h0074191

tests by race: A critical analysis of three recent studies. *Journal of Applied Psychology,
63*(1), 1-11. doi: 10.1037/0021-9010.63.1.1


IHS Inc. (2015). Hispanics will account for more than 40 percent of U.S. employment in the next


Applied Psychology, 70*(1), 148-156. doi: 10.1037/0021-9010.70.1.148


