Standardizing Biases:
Selection Devices and the Quantification of Race

Daniel Hirschman
Brown University
daniel_hirschman@brown.edu

Emily Adlin Bosk
Rutgers University
emily.bosk@ssw.rutgers.edu

Abstract:

Racial inequality persists despite major advances in formal, legal equality. Scholars and policymakers argue that individual biases (both explicit and implicit) combine with subjective organizational decision-making practices to perpetuate racial inequality. The standardization of decision-making offers a potential solution, promising to eliminate the subjectivity that biases consequential decisions. We ask, under what conditions may standardization reduce racial inequality? Drawing on research in science studies and law and society, we argue that standardization must be understood as a heterogeneous practice capable of producing very different outcomes depending on the details of the standard and the organizational infrastructure surrounding its use. We compare "selection devices" – simple quantified tools for making allocation decisions – in undergraduate admissions and child welfare to highlight the complex relationships between race and standardization. Child welfare agencies adopted a colorblind actuarial device that attempted to predict which children were most at risk, and then make decisions based on those predictions. In contrast, the University of Michigan's points system explicitly considered and valued race, with the goal of increasing minority student enrollments in the context of promoting student body diversity. Comparing these cases demonstrates how actuarial standardization practices, including those adopted with the intention of reducing racial inequality, tend to reinforce an unequal status quo by ideologically reconfiguring mutable social structures into immutable individual risk factors. In contrast, non-actuarial practices that explicitly promote racial equality are vulnerable to political challenges, as they violate norms of colorblindness and cannot be defended in terms of their predictive validity.

Introduction

The persistence of racial inequality in the post-Jim Crow United States presents an enduring challenge. The Civil Rights revolution of the mid-20th century produced major legal reforms, including federal bans on racial discrimination in employment, education, housing, and credit. Despite gains in formal equality, racial inequality persists in each of those same domains – employment (Pager et al 2009), education (Lewis and Diamond 2015), housing (Rugh and Massey 2014), credit (Hwang et al 2015), and others, including child welfare (Roberts 2003, Putnam-Hornstein et al 2013) and criminal justice (Pettit and Western 2004).

Many scholars, policymakers, and activists have pointed to *standardization* as a technique for combating persistent racial inequality. By standardization, we refer to attempts to formalize organizational practices to reduce the variability of decision-making. Advocates of standardization argue that one mechanism for persistent racial inequality is racially-biased decisions resulting from either explicit prejudice, understood as on the decline (Manning et al 2015), or implicit bias, understood as pervasive (Quillian 2006). When individuals are provided with little formalized means for making decisions, racially-biased determinations will often insert themselves. In contrast, the more standardized a system is, the less it should matter *who* makes a particular decision. Reskin (2012: 31) puts the argument cleanly: "Because discretionary decisions are necessarily subjective, they open the door to cognitive (as well as intentional) bias." Removing discretion, and replacing it with standardization, thus promises to ameliorate racial inequality by limiting the potential for bias to infiltrate the decision-making process. This article examines the promise and limitation of standardization to transform decision-making as a means for reducing biased decisions and unequal outcomes.

While advocates of standardization typically identify racially-disproportionate outcomes as resulting from some combination of explicit prejudice and implicit bias, sociologists of race have long argued that both historical and contemporary racial inequality must be understood, instead, as structural and systemic (Bonilla-Silva 1997, Golash-Boza 2016). Sociologists of race and organizations have also demonstrated how organizational practices, even those that are seemingly race-neutral, may produce and reproduce racial inequality independently of the particular biases or prejudices of individual actors (Wooten and Coulette 2017, Ray 2019).

Understanding racism as a structural phenomenon, *we ask: how and under what conditions can standardization reduce racial inequality?* What are the limits on the promises of standardization to transform unequal outcomes? We examine the tension between the promise of standardization to solve issues of subjectivity that generate unequal outcomes and the potential to encode structural inequality in the context of a relatively streamlined form of standardized decision-making: *selection devices*. Selection devices are simple standardization tools designed for making categorical decisions about individuals. Hailed as an antidote to implicit bias, selection devices have also been denounced as another mechanism for reproducing categorical inequality in the United States. While proponents and critics tend to assume that selection devices work more or less as intended, the sociology of standardization reveals the social processes embedded within attempts to create uniform outcomes and the attendant resulting variability. Put another way, standards are not themselves standardized.

In this article, we compare across two selection devices in order to better understand the relationship between race and various forms of standardized decision-making. Our cases examine

selection devices in child welfare and college admissions, which themselves exemplify the two distinct diagnoses of mechanisms for racial inequality (psychological vs structural). In both cases, the introduction of standardization was intended (in part) to reduce racial inequality. In the first case, Child Protective Service (CPS) agencies adopted a colorblind device designed to improve consistency in decision-making and eliminate implicit racial bias in decision-making in order to reduce racial disproportionality. In the second case, admissions officers at the University of Michigan turned to a race-conscious device in order to help create a more racially diverse class in light of the systemic disadvantages faced by minority applicants. We draw on insights from the sociology of race, law and society, and science studies to make sense of these devices and the controversies in which they became embroiled. Our analysis yields three main findings about the limits of standardization to transform social hierarchies.

First, there is no single relationship between standardization and race. Standardization and objectivity are often conflated concepts in discussions about the need for more mechanical forms of decision-making, and both are often conflated with quantification. Quantification seems to provide a rational, objective tool for standardizing decisions (Porter 1995). But there is more than one way to use numbers to make a given decision, numbers are not value free, and not all numbers carry the same air of objectivity and rationality (Hirschman et al 2016). Racism is itself easy to standardize. Even avowedly antiracist standardization initiatives approach the task very differently. As the examples of affirmative action and child welfare show, similar intentions can lead to very different solutions, which in turn yield different outcomes.

Second, we assert that actuarialism, the most prominent and lauded form of standardization in many domains, is in tension with decision-making intended to reduce racial disparities. Actuarial methods create a simple, quantitative system for making decisions previously entrusted to expert judgment. This method involves predicting future outcomes based on past experiences. By redefining fairness in terms of riskiness – being fair to someone means treating them *according to the prediction of their risk* – actuarial techniques tend to reinforce structural disadvantage (Simon 1988). Actuarialism transforms risk into an immutable characteristic of an individual instead of a potentially mutable characteristic of that individual's social context (Starr 2014). In a context of extreme racial inequality, actuarialism identifies and then builds on structural disadvantages that cannot be solved through interventions aimed at decisionmakers alone. Child welfare agencies adopted actuarial devices, but in this context, actuarial standardization may fail to reduce racial disproportionality in outcomes because this disproportionality likely does not result primarily from individual-level bias, but rather from structural inequalities including racialized disparities in poverty.

Third, non-actuarial standardization *may* produce decision-making practices that successfully reduce racial disparities, but may also lose the appearance of objectivity that insulates controversial decisions from challenges. Affirmative action in undergraduate admissions offers one example where selection devices were not developed solely as a more accurate classification system; instead those devices *directly* valued equalizing racial outcomes in the service of promoting racial diversity. These devices proved politically weak, however, as successful legal challenges forced their abandonment. If actuarial technique are strong because they scramble politically salient social identities and justify themselves based on the quality of their accuracy then non-actuarial forms of quantification may be weak if they *fail* to scramble those identities, and fail to justify themselves in terms of predictive validity.

The remainder of this article proceeds as follows. First, we synthesize existing conversations on race, standards and decision-making to establish a framework for understanding our cases. Second, we describe our methods and data. We then turn to the cases proper. For each case, we explore the motivations behind standardizing decision making, the actual selection devices used for making decisions, the success or failure of the device in reducing racial inequality, and the challenges that emerged to the device's implementation.

Theorizing Race, Theorizing Standards

In this section, we first synthesize key arguments from the sociology of race before turning to research on standards and quantification. We then combine these arguments to assess existing claims about the role of standardized organizational practices in reducing racial inequality.

The Changing Structures of Racism

Drawing on the sociology of race, and especially race and organizations, we situate our research within the following premises. First, race and racism must be understood as structural, systemic phenomena not reducible to individual psychology. Second, organizations in particular must be understood as racialized. Third, colorblind racism and diversity have replaced explicit racism as key ideological supports for the persistence of racial inequality in organizations.

During the 20th century, structural understandings of racism gave way to psychological understandings in public and political discourse (Gordon 2015). Sociologists of race, in contrast, have long advocated a structural understanding of race and racism (Golash-Boza 2016). Bonilla-Silva (1997) argues that American society should be understood through the lens of racialized social systems, and that race and racism are "normal", constitutive parts of American economic, political, and social institutions, rather than an unusual feature of individual ideology. Racism in this perspective is "the organizational map that guides actions of racial actors in society. It becomes as real as the racial relations it organizes." (Bonilla-Silva 2001: 44).

While much of the sociology of race has focused on micro- and macrolevel practices that perpetuate racism, a recent strand of research has emphasized the importance of meso-level practices situated inside formal organizations (Wooten and Coulette 2017, Ray 2019). Wooten and Coulette (2017) identify how organizations themselves may have racialized identities and reputations, and how those identities reflect the unequal distribution of resources. Ray (2019), demonstrates how organizations empower or disempower individuals and legitimate racialized inequalities. Taken together, this work highlights how much of the production and reproduction of racial inequality happens at the interface between individuals and organizations.

The persistence of racial inequality alongside gains in formal equality has spurred the creation of new explanatory ideologies. At the individual level, theories of implicit bias or unconscious racism are most prominent (Quillian 2006). At the organizational level, diversity came to be the language through which organizations could talk about race as one of many benign differences (Berrey 2015).

These discourses of unconscious racism and diversity have strongly shaped organizational responses to persistent racial inequality. Two of the most common reactions to allegations of racism are implicit bias trainings and diversity trainings. Beyond these trainings, a psychological

understanding of racism has led to a call for organizational solutions that focus on removing implicitly (and explicitly) racist subjectivities from consequential decision-making practices. Scholars argue that eliminating discretion in contexts like mortgage lending and prosecutorial decisions has reduced racial discrepancies, and suggests that such processes could work in other institutional domains (Reskin 2012). Similarly, researchers studying child welfare have proposed standardizing investigator decisions as a means to ameliorate persistent racial disproportionality (see Bosk 2015 for a review).

In contrast, sociologists of race have identified both implicit bias and diversity as pillars of colorblind racism (Bonilla-Silva 2017, Kahn 2017) and argued that implicit bias and diversity trainings are ineffective for the larger project of reducing racial inequality (Russell-Brown 2018, Kalev et al 2006). Thus far, sociologists of race have attended less to the consequences of standardizing decision-making for racial inequality. In the following section, we situate these proposed organizational decision-making practices within conversations about the role of standards, quantification, and actuarial practices.

Standardization, Quantification, Actuarialism

Timmermans and Epstein (2010: 71) define standardization as "a process of constructing uniformities across time and space, through the generation of agreed-upon rules." The rules, or standards, generated by these processes knit together distant actors and create coherence that makes possible large-scale, complex organizations. Actors pursue standardization projects for diverse reasons, including efficiency, effectiveness, regulatory compliance, and equality. We focus here on attempts to standardize decision-making, specifically categorical decisions about individual cases, and pay special attention to the ways that standards negotiate tensions surrounding race and racial inequality.

Organizations may employ qualitative or quantitative standards to equalize outcomes. Campaigns to "ban the box" asking applicants for employment about their criminal record are qualitative efforts to standardize the flow of information in the hiring process and increase the employment of ex-offenders. Tragically, "ban the box" campaigns may have had the unintended effect of *increasing* racial discrimination as employers instead rely on race as a proxy for criminality, exemplifying the potential unintended consequences of standardized decision-making for inequality (Agan & Starr 2018).

Organizations view quantification as one way to avoid such undesired outcomes. Quantification refers to the "the production and communication of numbers" (Espeland and Stevens 2008: 402). Quantified decision-making relies on numbers to inform or make decisions. Organizations often turn to numbers to bolster the objectivity of their decisions (Porter 1995). In particular, quantification provides "mechanical objectivity", a guarantee that decisions are free from individual biases, and thus defensible against outside criticism. Porter (1995) argues that quantification is a weapon of the weak, used by experts who have lost their authority to make unchallenged decisions. Yet scholarship has paid much less attention to the conditions under which quantification provides that authority (Hirschman et al 2016), particularly when related to such controversial subjects such as race and equality.

In order to produce effective, defensible, choices, many organizations turn to actuarial practices to quantify their decision-making. We define actuarial practices as decision-making systems that first

use statistical techniques to predict what factors lead to a successful outcome for an individual case (as defined by the organization), and then use those predictions to make decisions about individual cases (choosing cases in order to maximize the likelihood of success, or minimize the risk of failure). Inspired by the success of insurance pricing by actuaries, actuarial techniques are now widespread. Generally, these techniques use data on the actual outcomes of previous cases to determine what characteristics predict negative (or positive) outcomes. These characteristics are then transformed into a scoring system for classifying the likelihood of future outcomes, which in turn provides the basis for decisions.

Actuarialism, the ideological justification for actuarial techniques, brings with it a particular conception of fairness that equates fair treatment with outcomes determined by your predicted level of risk (Stone 1993). In credit scoring, actuarial fairness means that individuals at higher risk of default should either be denied credit or charged higher rates. In criminal justice, actuarial fairness implies that convicted criminals with a higher risk of reoffending should serve longer sentences.

The push for actuarial techniques as a replacement for individual judgment began as early as the first half of the 20th century, especially in the context of sentencing and parole (Harcourt 2006) and steadily gained ground. The movement began to gain ground in other domains with a line of papers beginning with Meehl (1954) and extending through Grove et al's (2000) meta-analysis of 136 studies of actuarial prediction techniques which determined that "on average, mechanical prediction techniques were about 10% more accurate than clinical predictions." (2000: 19). Notably, this broad push for statistical or actuarial decision-making was initially motivated in terms of the inconsistency (or low "reliability") of expert decisions and the relatively weak predictive power of decisions, not around issues of racial bias. More recent efforts, however, have framed these techniques as a tool for reducing racial disparities as in the case of child welfare (Bosk 2018). In the following section, we bring together critical insights on standardization and quantification with findings from the sociology of race to theorize the complex relationship between standardization and racial inequality and, more specifically, some of the pitfall of actuarial techniques.

Theorizing Standardization and Racial Inequality

We return now to our initial question: under what conditions can standardization reduce racial inequality? What are the conditions under which actuarial and non-actuarial forms of quantified decision-making succeed or fail in reducing racial inequality? Here, we connect our discussions of the complexities of standardization to the discussion of different diagnoses of persistent racial inequality.

Individualistic theories, like implicit bias, tend to explain persistent racial inequality as a function of decisions made by (intendedly) non-racist actors in the context of formally equal, race-neutral practices. In contrast, structural racism perspectives, including the racialized organizations approach, assert that racial inequality persists (in part) because formally equally and colorblind policies are racist in practice and effect.

The implicit bias perspective implies that standardization will be highly effective at reducing racial inequality as long as it succeeds at removing discretion from decision-makers. For example, Reskin (2012) cites research showing how discretionary decisions by real estate agents, mortgage

lenders, prosecutors, and public assistance workers all lead to discriminatory outcomes, but that the automation of mortgage lending has "reduced the racial gap" (Reskin 2012: 31) by removing discretion from the system.

The example of mortgage lending is a useful one for thinking through the promises and perils of standardization. As Reskin notes, automated mortgage lending based on credit scores has expanded access to mortgages. But those same actuarial technologies made possible the emergence of the subprime mortgage market and eventual housing crash (Poon 2009). Despite the seemingly automated, race-neutral character of the lending process, researchers have determined that racial minority borrowers were disproportionately tracked into subprime mortgages, even controlling for their credit scores (Rugh and Massey 2010; Hwang et al 2015). In a prescient 2005 article, Williams et al (2005) noted that changes in the mortgage industry had not eliminated racial inequality as much as reshaped it.

That standardized decision-making would lead to a transformation of inequality in mortgage lending rather than its elimination is unsurprising in light of scholarship on the complex and often contradictory consequences of standardization, and the scholarship on racialized organizations. When brought in to conversation, these perspectives suggest that seemingly race-neutral, objective standards may re-embed racial inequalities in new logics and justifications while simultaneously immunizing organizations from claims of explicit racial bias.

More broadly, scholarship on race and standardization suggests four problems that confront attempts to use actuarial techniques to address racial inequality: implementation, diagnosis, ideology, and politics. First, scholars in science studies and organizational sociology have argued that standards never work as advertised, and that organizational processes are hard to standardize (Timmermans and Epstein 2010). That is, the actual implementation of an actuarial technique may differ markedly from its design. In principle, at least, quantified decision-making completely removes subjectivity from the process of deciding about individual cases. In practice, judgment is (at best) moved up one level to the construction of the system.[1] In turn, the relevant potential biases transform from the individual decisionmaker to the system as a whole – what kind of data it considers, what factors are seen as relevant, how they are combined, and what goals the designers of the system choose to emphasize.

Second, because racial inequalities are best diagnosed as reflecting structural racism (Bonilla-Silva 1997), not (just) individual bias, efforts to reduce racial inequality through standardizing gatekeeping decision-making may have little effect. Third, law and society scholars have argued that actuarial decision-making may actually make discriminatory decisions harder to challenge. Actuarialism tends to bake inequality into the decision-making process, transmuting social disadvantages into seemingly objective measures of individual riskiness. Fourth, actuarial techniques are demobilizing, scrambling socially identities that might have served as the basis for political challenges (Simon 1988). Traditionally, political challenges to unjust decisions have involved mobilizing along recognizable social identities, especially race and gender. When feminists in the 1960s and 1970s mobilized to demand access to credit, they did so on the basis of unfair treatment against women as a group (Krippner 2017). The credit industry responded by dropping the explicit use of sex in credit decision-making, and turning to an actuarial system of credit scoring which relied on seemingly gender-neutral variables. This system still excluded many

---

[1] As we will discuss below, subjectivity frequently reemerges even at the level of the individual case.

women from access to credit, but rather than justifying the exclusion in terms of their sex, creditors pointed to the complex array of factors that predicted repayment, like length of credit history, that disfavored women. In this sense, actuarial techniques may prove difficult to contest not just because of the mechanical objectivity associated with quantification in general (Porter 1995), but also because their focus on risk factors undercuts mobilization based on socially meaningful categories.

In contrast, non-actuarial standards do not rely primarily on predictions of future outcomes to generate decisions, and thus may be more useful in combating racial inequality. In particular, non-actuarial systems present the opportunity to account for structural forms of inequality and to directly value equalizing outcomes based on race. But for the same reason that actuarial standards are difficult to challenge, non-actuarial standards may be politically vulnerable. Without the justification of statistically superior decision-making according to a mechanically objective metric, non-actuarial standards may be easier to challenge and more difficult to defend – especially when they explicitly challenge racially unequal outcomes. Our two cases exemplify these contrasting relationships between standardization and racial inequality: a politically defensible, but largely ineffective, actuarial standard in child welfare compared to a politically vulnerable, but promising, standard in undergraduate admissions.

Cases & Methods

In the remainder of our analysis, we examine these tensions of standardization and racial inequality in the contexts of child welfare and undergraduate admissions in order to understand the potential and limitations of standardization to transform social hierarchies. These two cases share several important features, including the explicit goal of reducing racial inequality, and the form of quantified decision-making employed by the relevant organizations. Borrowing a term from Duncan et al, we identify this form as a "selection device": "an apparatus consisting of a selection instrument (e.g. a test) which, independently of the judgment of the administrator, assigns a score to each applicant; and a rule for deciding which applicants shall be accepted and which ones rejected, solely on the basis of their assigned scores." (Duncan et al 1953: 573). In child welfare, a non-profit organization designed a selection device to determine how a case should be approached by state child welfare agencies after a substantiated charge of neglect or abuse, including whether a child should remain at home. In undergraduate admissions, the University of Michigan developed a selection device to determine whether to admit applicants to its undergraduate program.

The cases differ in key respects as well, including most notably their respective diagnoses of the cause of racial inequality and in their use of actuarial techniques. Child welfare agencies adopted a formally race-neutral selection device statistically designed to classify cases according to the likelihood of future abuse and neglect with the theory that this device would improve decision-making and reduce racial inequality by constraining the subjectivity of decisionmakers. In contrast, the selection device developed at the University of Michigan was not designed to classify which students would be the most likely to succeed; rather, the device balanced several competing concerns, including the need to create a racially diverse student body, and did so by explicitly considering race.

Our data come from primary and secondary research. The second author conducted extensive primary data collection on child welfare decision-making in two states, including 80 interviews

with child welfare workers and agency administrators along with secondary research on decision-making tools and bureaucratic procedures. The first author, as part of a collaborative research project, conducted primary historical research on college admissions, including analyzing internal admissions documents from University of Michigan's archives, along with court documents and media coverage (see Author Citations for details). For each case, we analyze the development of the selection device, the intentions of the developers and implementers, the functioning of the device itself, the challenges raised against the new quantified decision-making system, and the system's overall performance in terms of reducing racial inequality. We summarize our analysis in Table 1.

Table 1: Case Comparison

	Child Welfare	Undergraduate Admissions
Decision	Level of intervention and services, including whether a child should be removed from home.	Admit or reject undergraduate applicant.
Selection Device	SDM Risk Assessment	Points System
Motivations	Reduce racial disproportionality by eliminating implicit bias; improve decision-making through better use of evidence and increased consistency.	Increase diversity by increasing minority student enrollments, decrease costs of processing applications.
Actuarial?	Yes, designed to predict the likelihood of future abuse or neglect.	No, designed to reproduce previous admissions decisions reflecting a mix of priorities.
Use of Race	Not explicitly included in system.	Explicitly included as a separate factor.
Change in Racial Inequality	None.	Significant increase in minority student enrollment.
Political Pushback	None.	System challenged, ruled unconstitutional, and eliminated.

Diverse Standards

In the following sections, we describe the development of selection devices in child welfare and undergraduate admissions. We begin with child welfare to showcase how actuarial standardization succeeds at achieving mechanical objectivity, but fails to address racial disproportionality. We then turn to undergraduate admissions at the University of Michigan to showcase how non-actuarial quantified decision-making that makes explicit use of race as a variable can succeed in promoting racial equality at the risk of provoking successful political backlash.

Structuring Child Welfare Decisions

Child welfare decisions are among the weightiest a state can make and are also associated with significant racial and ethnic disproportionality at every point in the process (from who is reported for child maltreatment to the number of children in foster care). While decisions about child protection are rooted in child safety mandates, sociological research makes clear that racialized and classed narratives about families, motherhood, and appropriate caregiving shape this process in ways that can punish minority families (Reich 2005; Roberts 2003).

The move to standardized, actuarial decision-making in child welfare is a relatively recent phenomenon. From the turn of the twentieth century until the mid 1980s, clinical and bureaucratic decision-making models were the only available tools used to frame decisions regarding child welfare substantiations and placements (e.g. whether the allegation of maltreatment is found to be true and whether a child is safe to remain at home). Prior to the 1980s, debates about decision-making centered around the best ways to make clinical decisions, the impact of race and gender on the decision-making process and its outcomes, and the extent to which bureaucratic policies should influence clinical decisions.

In the mid-1980s, the Children's Research Center (CRC) became the first organization to develop actuarial-based risk assessments for child welfare. Figure 1 is an example of the risk assessment designed by the CRC for the State of Michigan in 1998 to predict the risk of future maltreatment. This model is a relatively simple tool that classifies cases according to the likelihood of future instances of abuse or neglect using variables associated with re-entry into the child welfare system in past cases.

RISK ASSESSMENT OF ABUSE/NEGLECT
Michigan Department of Human Services

Caretaker Name: _____   PS Case Number: _____

Load Number: _____   Assessment Date: _____

NEGLECT	Score
N1. Current complaint and/or finding includes neglect	
a. No	0
b. Yes	2 ___
N2. Number of prior assigned neglect complaints and/or findings	
a. One or less	0
b. Two or more	2 ___
N3. Number of children in the household	
a. Three or less	0
b. Four or more	1 ___
N4. Primary caretaker's social support	
a. Appropriate and available social support	0
b. Limited or negative support *(check all that apply)*	
___ No or limited supportive relationships with relatives/friends/neighbors	1
___ Relatives/friends/neighbors have negative impact	1 ___
N5. Primary caretaker is unable/unwilling to control impulses	
a. No	0
b. Yes	1 ___
N6. Primary caretaker provides inadequate physical care and/or inadequate supervision of child(ren)	
a. No	0
b. Yes *(check all that apply)*	
___ Provides inadequate physical care	1
___ Provides inadequate supervision	1 ___
N7. Primary caretaker currently has a mental health problem	
a. No	0
b. Yes	1 ___
N8. Primary caretaker involved in harmful relationships	
a. No	0
b. Harmful relationship(s) or one domestic violence incident	1
c. Multiple (two or more) domestic violence incidents	2 ___
N9. Primary caretaker currently has a substance abuse problem	
a. No	0
b. Yes	1 ___
N10. Family is homeless or children are unsafe due to housing conditions	
a. No	0
b. Yes *(check all that apply)*	
___ Family is homeless or about to be evicted	2
___ Housing is physically unsafe	2 ___
N11. Primary caretaker able to put child needs ahead of own	
a. Yes	0
b. No	1 ___

ABUSE	Score
A1. Current complaint and/or finding includes mental injury	
a. No	0
b. Yes	1 ___
A2. Number of prior assigned abuse complaints and/or findings	
a. None	-1
b. One or two	0
c. Three or more	1 ___
A3. Age of youngest child	
a. Seven years or older	0
b. Six years or younger	1 ___
A4. Number of children in the household	
a. Two or less	0
b. Three or more	2 ___
A5. Either caretaker was abused and/or neglected as a child	
a. No	0
b. Yes	2 ___
A6. Secondary caretaker has low self-esteem	
a. ___ No secondary caretaker *(check if applicable)*	0
___ No	0
b. ___ Yes	1 ___
A7. Either caretaker is domineering and/or employs excessive and/or inappropriate discipline	
a. No, neither caretaker	0
b. Yes *(check all that apply)*	
___ Domineering	1
___ Inappropriate discipline	1 ___
A8. Either caretaker has current or a history of domestic violence	
a. No, neither caretaker	0
b. Yes	1 ___
A9. A child in the household has one or more of the following characteristics	
a. No child has any of the characteristics below	0
b. Yes *(check all that apply and indicate highest score)*	
___ Developmental disability	1
___ History of delinquency	1
___ Mental health issue	2
___ Behavioral issues	2 ___
A10. All caretakers are motivated to improve parenting skills	
a. Yes, all caretakers motivated or improvement not needed	-1
b. Yes, caretakers are willing to participate	-1
c. No, one or both caretakers needs to improve parenting skills but will not participate	0 ___
A11. Primary caretaker views incident less seriously than department	
a. No	0
b. Yes, views incident less seriously	1 ___

Figure 1: Selection Device for Making Child Welfare Decisions

The Risk Assessment (RA) contains two sets of inventories, one for neglect and one for abuse. Factors associated with abuse and neglect like parental history of abuse, history or presence of domestic violence, number of children in the house, age of caregiver, and previous involvement with CPS, make up the scales. Points are assigned to each item that the caseworker endorses. For example, if a worker agreed with the statement that the "parent has a mental health problem," one point would be assigned.

Point values for each statement are weighted to reflect distinct levels of severity as determined by statistical regression models. The total points from the neglect inventory and the abuse inventory are calculated separately. The highest number from either inventory then sorts a family into Intensive, High, Moderate, or Low risk for the recurrence of maltreatment. The goal of the RA is to accurately classify which children are most at risk for re-abuse and to direct appropriate

intensive services to those at highest risk. The developers assert that the RA objectively classifies cases based on the likelihood that a caregiver will maltreat their child(ren) in the future. In promotional materials the Children's Research Center states: "It is important that if given the same information, all workers, despite the personal biases they may have, consistently evaluate families in an objective way.[2]

In the late 1990s, the CRC funded validation studies conducted on early adopter states which concluded that actuarial prediction was the most valid and reliable method for child welfare decision-making (Baird and Wagner 2000). Although other forms of quantified decision-making tools were being tested and used at this time, these validation studies and other comparative research made a strong case for the adoption of actuarial-based tools (Baird and Wagner 2000; Baird et al. 1999). Building on this research, CRC expanded the use of actuarial assessments to other decision points in the child welfare decision-making process such as reunification (CRC 2008). Bringing a series of actuarial and other clinically structured standardized tools under the umbrella of one, easy to use, step by step procedure, the Structured Decision Making Model (SDM) provided a process for case decision-making that mixed mechanical with clinical judgment. The RA represents the most prominent component of the SDM and is the focus of evaluations of this model. By the end of the 1990s, 10 states adopted the SDM for guiding child welfare decision-making and today over 33 states use the SDM.

As the SDM was being adopted, the field of psychology began to emphasize the superiority of actuarial prediction over clinical assessment more generally. As noted above, Dawes, Faust, and Meehl's (1989) meta-analysis argued that actuarial-based methods were more accurate than clinical predictions. Although Dawes et al (1989) acknowledged that actuarial-based models were not superior in all cases, their study was framed by the CRC as settling the issue over what method to use to approach cases. In promotional material, the CRC prominently quoted Dawes stating:

> In the last 50 years or so, the question of whether a statistical or clinical approach is superior has been the subject of extensive empirical investigation; statistical vs. clinical methods of predicting important human outcomes have been compared with each other, in what might be described as a "contest." The results have been uniform. Even fairly simple statistical models outperform clinical judgment. (Dawes, 1993, in CRC, 1999)

Currently, the CRC asserts on its website that SDM is much more accurate than any other form of decision-making noting: "The SDM model offers an elegant and comprehensive way to incorporate research and consistency into key child welfare decisions. To date, no set of CPS assessments has demonstrated the degree of reliability and validity, nor the improved outcomes of the SDM Model".[3]

The shift toward actuarial-based risk assessment in child welfare emerged at the intersection of three interrelated concerns: the low reliability of clinical decision-making, the emergence of the Evidence-Based Medicine movement, and the need to address the disproportionality of ethnic and racial minorities in the child welfare system. These are discussed in order.

---

[2] https://www.nccdglobal.org/assessment/sdm-structured-decision-making-systems/child-welfare, accessed 2012.
[3] http://www.nccdglobal.org/assessment/sdm-structured-decision-making-systems/child-welfare, accessed 12/17/18. Because the only studies evaluating the SDM have focused on validating the RA (Schwalbe 2009), we assume that it is this tool (and another actuarial-based assessment for reunification) that are the focus of this quote.

Reliability issues with clinical decision-making became especially salient following research conducted in the late 1980s and 1990s that suggested that case decisions were highly dependent on which worker a family received and the larger institutional context in which the case was processed, and less dependent on the seemingly objective facts of the case (Stein and Rzepnicki 1983; Rossi et al 1999; Lindsey 1992). Further, research demonstrated little connection between the criteria CPS workers used to make a judgment and established risk factors related to re-abuse (Baird and Wagner 2000; Gambrill and Shlonsky 2001). Compounding these issues in decision-making were the difficult conditions under which CPS workers made judgments, namely with limited information, under time pressure, and in conditions of high uncertainty (Gambrill and Shlonsky 2001). Since decision-making errors can lead to the death or the unnecessary separation of a child from her family, improving reliability in child welfare decision-making was understood to be essential.

Moves toward Evidence-Based Medicine in both health care and psychology in the late 1990s and early 2000s also heavily influenced the expansion of the SDM into state child welfare policy. Concerned with improving the effectiveness of social work, resolving continued gaps between research and practice, and aligning professional standards with those in medicine, social work researchers and practitioners have advocated for the necessity of Evidence-Based Practice (EBP) as a means to ensure the scientific legitimacy of the field and provide interventions that "work," leading EBP to become the guiding model for theory and practice in the field (Okpych and Yu 2014).

The emergence of the EBP movement coincided with the introduction of new federal policies such as the Government Performance and Results Act of 1993 which mandated the collection of data both on performance targets and outcomes for major federal programs (Hatry 1997, Okypych and Yu 2014). In child welfare, between 2000-2004, all states were required to conduct Child and Family Service Reviews (CFSR), which included new substantiations after the close of an investigation as a variable for state performance. This criterion further generated interest in "simple risk assessment procedures with known predictive attributes" regarding the risk of future maltreatment" (Shlonsky and Wagner 2005: 412).

In response, the CRC framed the SDM not only as the most accurate form of decision-making in child welfare, but also a model that facilitates compliance with the new guidelines. Use of the SDM expanded to over 30 states by 2010 and the CRC began exporting their model internationally. According to the CRC, the SDM is "now the most widely used case management model in the United States" (CRC 2008). Research in child welfare considers actuarial-based risk assessment to be the "gold standard" for case decision-making (Schwalbe 2009: 205).

Finally, and most important to our larger argument, the significant overrepresentation of ethnic and racial minorities at every point in the child welfare system including, reports, placement in foster care, time in foster care, and family reunification (Sedlak & Broadhurst 1996) also influenced the development and expansion of actuarial based methods for decision-making in child welfare. While disproportionality in the child welfare system is currently understood to be a complex issue with multiple causes (Putnam-Hornstein et al 2013, Pelton, 2015, Font et al 2015), child welfare research in the 1990s framed the persistent overrepresentation of racial and ethnic minorities as partly a result of decisionmaker bias in which "race, rather than risk, was the relevant factor" (Putnam-Hornstein et al 2013: 34). Decreasing cognitive errors in decision-making that

rested on the heuristic errors and implicit biases of the child welfare worker therefore, represented an important policy corrective to race-based differential treatment (Gambrill and Shlonsky 2000).

The CRC promoted the RA as a tool to advance this goal (CRC 2008, CRC 1999). On an earlier version of their website, the CRC listed equity alongside reliability, validity, and utility as forming the basis on which the value of the SDM rests. Describing equity, the page noted: "SDM assessments ensure that critical case characteristics, risk factors, and domains of functioning are assessed for every client, every time, regardless of social differences. Detailed definitions for assessment items increase the likelihood that workers assess all clients using a common framework."[4] Through the formulation of standard criteria for evaluation and their systematic application, the CRC promised a mechanism to reduce racial disproportionality.

In their promotional materials from the 2000s, the CRC continued to assert that actuarial-based decision-making could equalize racial outcomes in child welfare decision-making:

> In all CRC studies to date, the developed risk systems *also promote equity in decision making.* Because equity is a major principle of the development process, the proportion of various races and ethnic groups assigned to each risk level is virtually identical in all jurisdictions. These results suggest that well-structured assessment tools and decision making systems *can help overcome some of the racial disparities resulting from traditional practices* (CRC 2008: 12 emphasis added)

Disentangling race and ethnicity from the decision-making process was thus one explicit goal of the developers of the SDM. The list of variables collected does not include race or ethnicity. By explicitly leaving race out of the model, by focusing on individual behaviors and family circumstances, and by standardizing the content CPS workers use to make decisions, the SDM understands itself to be removing race from the decision-making process.

Despite these intentions, proxies for disadvantage slip in through questions about residential status (whether a family has a home), employment, and number of children on the neglect inventory. And much as in the case of parole predictions (Harcourt 2015), the SDM includes past complaints and offenses, a variable that is far from race-neutral. Research on racial disparities in child welfare highlights that racial minorities are overrepresented in allegations of maltreatment to CPS for all groups of reporters (Krase 2013). By including criteria associated with disproportionality such as past complaints, the SDM (to at least some extent) reinforces the problem it was designed to correct.

Further, new research on racial disproportionality in the child welfare system highlights the role that socio-economic status (SES) plays in driving disproportionate outcomes (Pelton 2015, Putnam-Hornstein et al 2013, Drake et al 2011, Font et al 2012). This line of work suggests that when economic disadvantage is taken in to account, race-based differences in maltreatment rates are either reduced, eliminated, or reversed. Racial disproportionality in the child welfare system therefore may not be mostly driven by bias in the decisions of CPS workers, but instead may reflect structurally situated socio-economic risk factors. Proxies for race in the RA such as residential

---

[4] http://www.nccdglobal.org/assessment/sdm-structured-decision-making-systems/child-welfare, 2011.

status and employment are also proxies for SES, continuing to confound attempts to reduce race-based differential results.

Beyond problematic proxy variables, the SDM also includes many variables that require substantial judgment on the part of social workers to code, and thus reintroduce the possibility for cognitive racial prejudices to influence decisions. As Bosk (2018) shows, even a question as straightforward as "Primary caretaker currently has a substance abuse problem" requires interpretive acts on the part of caseworkers – how should one score a parent with a history of alcoholism but who is currently sober and regularly attends AA? Bosk's study, examining how child welfare workers utilize and make meaning of actuarial-based risk assessments in practice, demonstrates that workers have a wide range of responses and approaches to them. While a minority of workers understand the RA to be an important tool to eliminate bias and disparate assessments of similar cases, the majority of child welfare workers view the RA as one more way to encode subjectivity in to the decision-making process (Bosk 2018). Workers' critiques focus on the ways that, in their view, the RA penalizes caregivers for being victims of circumstance, situating risk in factors that are not changeable or overtly related to behavior such as the number of children in the home, the age of the child, and whether or not the caregiver has a history of experiencing intimate partner violence. Workers lack of confidence in the RA to objectively or fairly assess risk, which can lead them to ignore it in their decision-making (Bosk 2018).

We could locate only two specific independent evaluations of racial disparities after the introduction of the SDM. These analyses found that while the introduction of the SDM was more successful in accurately classifying low, moderate, and high-risk cases, it did not reduce racial disproportionality (Osterling et al 2008, Miller 2011). Further, at a national level, racial and ethnic disproportionality at all points in the child welfare system remains a persistent and major cause for concern (Putnam-Hornstein et al 2013, Font 2013), suggesting that the widespread take-up of the SDM has had little impact on overall racial disparities. These results are consistent with newer studies which find that racial disproportionalities in child welfare largely reflect structural inequalities, particularly those related to economic inequality and thus that even when implicit bias strongly shapes child welfare decisions, standardization efforts like the RA will have limited effects in reducing racial disproportionality.

Affirmative Action in College Admissions

College admissions have been a site of controversy since the early 20th century (Karabel 2005). Access to selective colleges has been understood as a key path to upward class mobility or upper-class reproduction. For colleges, admissions practices have served multiple functions, including legitimacy, prestige, and financial stability (Killgore 2009). In the first half of the 20th century, selective colleges turned towards quantified assessments of academic merit in order to identify promising (white) students from less affluent backgrounds (Lemann 1999). Inspired by the civil rights movements of the early 1960s, some selective colleges also began to engage in programs of affirmative action to increase the enrollment of non-white students (Stulberg and Chen 2014). These practices spread such that by the mid-1990s approximately 60% of selective colleges claimed to consider race in admissions (Hirschman and Berrey 2017).

Our research here focuses on a single case within higher education, the University of Michigan. In 2003, the Supreme Court ruled unconstitutional Michigan's use of a particular selection device, the famous "points system," because it mechanically awarded 20 points to underrepresented

minority applicants (out of 100 needed for admission), while simultaneously upholding the constitutionality of more holistic modes of affirmative action. We briefly review the history of affirmative action at Michigan prior to these decisions and use debates over the points system to showcase how the University arrived at a non-actuarial selection device that successfully streamlined and standardized the admissions processes while increasing minority enrollments, but which failed to survive political challenges.

Like some peer institutions, Michigan began pursuing a policy of intentionally recruiting minority (especially Black) students in the mid-1960s (Stulberg and Chen 2014). At the time, Black students represented just .1% of the student body. Initial efforts to increase minority recruitment revolved around an alternative path to admissions and financial aid for minority students identified as promising. These efforts were initially justified in terms of a logic of redistribution and creating opportunities for previously marginalized groups (Berrey 2015). In the 1960s and 1970s, Michigan expanded its affirmative action programs and moved away from a model based on personal connections between admissions officers and potential minority applicants in order to streamline the process and further increase minority enrollments. These efforts were undertaken in part in response to the Black Action Movement, which in 1969 demanded that Michigan increase minority student enrollment to 10% (at the time, it was 3.4%) (Brune 1994).

In 1978, the legal context for affirmative action changed. In *Bakke v. University of California Davis*, the Supreme Court ruled that quotas were unconstitutional violations of equal protection, but the somewhat opaque ruling left open the possibility that some forms of affirmative action could be justified. According to Justice Powell's decision, universities might have a constitutionally protected interest in promoting diversity that justified affirmative action even as the Court rejected arguments based on reducing societal inequalities (Berrey 2015).

The University of Michigan revamped its admissions practices with this ruling in mind. But while many selective universities turned to individualized review, Michigan began creating a system of quantified admissions. Faced with increased application volume and tight budgets, Michigan sought to enroll a diverse student body in as efficient a fashion as possible (Hirschman et al 2016). The system they created aimed to standardize race-conscious admissions without an explicit quota. In the 1980s and early 1990s, Michigan employed multiple "grids", which mapped a student's GPA and standardized test scores to an admission decision. Minority students were mapped onto separate grids, with lower thresholds for admission (Hirschman et al 2016: 280-282). These grids were made public following a 1995 FOIA request by an anti-affirmative action activist, who considered them "smoking gun" evidence of unconstitutional discrimination (Hirschman et al 2016: 284). Conservative legal advocates agreed, and chose Michigan for the lawsuits that became *Gratz v. Bollinger* and *Grutter v. Bollinger* in part because of the clarity of its quantified system of affirmative action.

In 1997, Michigan revised its grid system into a single page selection device known colloquially as the "points system." Figure 2 shows the points system as it was used from 1998 to 2003. It was this system which the Supreme Court would evaluate and eventually rule unconstitutional in *Gratz*.

This system differs in two important respects from the child welfare Risk Assessment. First, the points system was not developed in order to classify students based on past successful outcomes. That is, statisticians did not define some outcome (like graduation) and then see what variables best predicted that outcome. Instead, statisticians built a system that incorporated heterogeneous

organizational imperatives: to build a diverse class along racial and gender lines (e.g. 5 points for men in nursing), to promote alumni relationships (4 points for legacies), and to serve residents of the state (10 points for Michigan residents).[5]

Place student label here or fill in the following information   Confident

Name _____

SIN _____

Unit _____

Term _____

Selection Index Worksheet

(For LS&A and selected Schools and Divisions, other than Engineeri.

For Internal Use Only

Points (circle)	ACADEMIC	
40	2.0	
42	2.1	
44	2.2	
46	2.3	
48	2.4	
50	2.5	
52	2.6	
54	2.7	
56	2.8	
GPA 58	2.9	
60	3.0	
62	3.1	
64	3.2	
66	3.3	
68	3.4	
70	3.5	
72	3.6	
74	3.7	
76	3.8	
78	3.9	
80	4.0	

School Factor		
0	0	
2	1	
4	2	
6	3	
8	4	
10	5	

Curricu- lum Factor		
-4	-2	
-2	-1	
0	0	
2	1	
4	2	
6	3	
8	4	

Points (circle one)	TEST SCORE	
ACT/ SAT I		
0	01-19	400-920
6	20-21	930-1000
10	22-26	1010-1190
11	27-30	1200-1350
12	31-36	1360-1600

OTHER FACTORS

(Maximum 40 points)

Points (circle)

Geography

10   Michigan Resident

6   Underrepresented Michigan County

2   Underrepresented State

Alumni

4   Assign only 1 option   Legacy (parents/stepparents)

     or

1   Other (grandparents, siblings, spouses)

Essay

1   Outstanding Essay

Personal Achievement

1   Assign only 1 option   State

3   Regional

5   National

Leadership & Service

1   Assign only 1 option   State

3   Regional

5   National

Miscellaneous

20   Socio-economic Disadvantage

20   Underrepresented Racial/Ethnic Minority Identification or Educatic

5   Assign only 1 option   Men in Nursing

20   Scholarship Athlete (assigned by athletic counselor only)

20   Provost's Discretion

Add points in this column for sub score 2

UMA 005893

TOTALS

Sub score 1 + Sub score 2 = Selection Index

Figure 2: A selection device for undergraduate admissions at the University of Michigan.

---

[5] See the deposition of David Hunter, Admissions Lawsuit Collection, Box 14, file: "Defendant Motions," Bentley Historical Library, University of Michigan.

Second, and most controversially, the points system explicitly considered the race of applicants. Underrepresented minority students (defined as Black, Hispanic, or Native American) received 20 points. Making race explicit also *separated* race from other variables that may have a racialized component – the SAT may have a racial bias, underrepresented minority students may attend less well-resourced schools, but race here is treated as an isolated variable.

This system (along with other efforts to recruit a diverse student body) yielded a significant increase in racial diversity. By 2002, Michigan had almost 10% Black students in its entering class.

In 2003, the Supreme Court scrutinized the points system in *Gratz v. Bollinger*. In oral arguments, the Justices debated whether 20 points for minority applicants was appropriate – was the equivalent of a full point of GPA too much? Was it fair that students from economically disadvantaged backgrounds received at most the same number of points as underrepresented minority students? The legible, transparent, and non-actuarially derived "20 points" was mentioned 19 times in oral arguments. For example, Breyer noted that a white Male athlete from a poor family would receive the same points as an African American from a rich family, and so the athlete "just can't overcome that 20 points – the best he can do is tie."[6]

These questions played into the eventual 5-4 decision ruling that the points system was not "narrowly tailored" to achieve the constitutionally protected interest of promoting diversity. The points system was not treated as a mechanically objective tool for making decisions (Porter 1995), but an ad hoc system that too coarsely advantaged minority students without clearly linking that advantage to the nebulous goal of promoting diversity. The *Gratz* majority opinion stated the "automatic distribution of 20 points has the effect of making 'the factor of race… decisive' for virtually every minimally qualified underrepresented minority applicant."[7] The points system, with its explicit and legible quantification of race, violated the compact binding together colorblindness and diversity (Moore and Bell 2011).

Guided by these two rulings, Michigan abandoned the points system and replaced it with a version of holistic assessment. This new system required a 60% increase in admissions staff (Hirschman et al 2016: 293). At the same time, minority enrollments plummeted; Black enrollment in 2004 was just 6%, down from almost 10% under the points system (New York Times 2013). The new system was both more expensive and less effective at yielding a racially diverse student body.

Discussion & Conclusion

Our empirical cases examined two intendedly antiracist selection devices to identify how actuarial and non-actuarial forms of standardization offer different affordances and constraints in the service of redressing racial inequality. Child welfare organizations identified individual, psychological sources of racial disproportionality and sought to address them with an actuarial tool designed to improve decision-making and increase objectivity. Its rollout had minimal effects on racial disproportionality in child welfare cases but faced virtually no political pushback. In contrast, at the University of Michigan, admissions staff produced a non-actuarial tool designed to facilitate

---

[6] *Gratz v. Bollinger* oral argument, available at https://www.oyez.org/cases/2002/02-516 .
[7] In contrast, in *Grutter*, the same Court ruled that the holistic review employed by Michigan's law school passed constitutional muster.

efficient decision-making. The points system recognized structural inequalities and directly valued increased racial minority enrollment in order to achieve its aims. This system was effective at increasing the enrollments of Black students to a level almost equal to the historical demands of Black student groups. But this system was also transparent, and transparently race-conscious, making it subject to significant challenge and as a violation of the paradoxical colorblind logic of diversity, which resulted in its elimination.

The contrasts between these two cases suggests three findings. First, there is no single relationship between standardization and race. The SDM and the points system reflected different diagnoses of the causes of racial inequality and offered different solutions. Both were forms of standardization intended to reduce racial inequality, but their construction and effects were quite different.

Second, actuarial techniques exist in tension with efforts to reduce racial inequality. The SDM and its Risk Assessment tool likely failed to reduce disproportionality, in part, *because* it is based on actuarial techniques. Despite not including race as an explicit variable, the RA likely detected extreme structural racial inequalities and transmuted them into risk classifications. These classifications, in turn, became seemingly race-neutral justifications for child welfare decisions. But nothing in this process could correct the disadvantages detected. Standards that rely on actuarial techniques are unlikely to succeed in promoting racial equality because of the way actuarial techniques tend to recast structural inequalities in terms of individual merit. Even if the Risk Assessment detects risk accurately, it cannot resolve the way these risks are disproportionately distributed.

Third, while actuarial standards may be limited in their ability to reduce racial inequality, non-actuarial standards may be vulnerable to challenges of bias precisely because they are not justified on a logic of prediction. The points system succeeded at helping the University of Michigan enroll more minority students, it failed to survive sustained political challenge. The University could not argue that the points system was statistically derived to identify the "best" students according to their likelihood of graduation or GPA or any other criteria. The system was not designed to classify the likelihood of success; it was designed to enroll a diverse class. Non-actuarial techniques that transparently address racist structures are vulnerable to political and legal challenges as they violate norms of colorblindness and the jurisprudence of diversity.

Together, these cases suggest a frustrating contradiction. To construct an effective anti-racist standard requires explicitly recognizing structural racial inequality. Yet constructing a politically defensible standard may require enacting a colorblind standard justified around individual understandings of racism. Unless and until political actors recognize structural racism, efforts to promote racial inequality through standardization are likely to fail.

References

Agan, Amanda and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *The Quarterly Journal of Economics* 133(1):191–235.

Baird, C, D Wagner, T Healy, and K Johnson. 1999. "Risk assessment in child protective services: consensus and actuarial model reliability." *Child Welfare*, 78: 723-48.

Baird, C, and D Wagner. 2000. "The relative validity of actuarial and consensus based risk assessment systems." *Children and youth services review*, 22: 839-71.

Berrey, Ellen. 2015. *The Enigma of Diversity*. University of Chicago Press.

Bonilla-Silva, Eduardo. 1997. "Rethinking Racism: Toward a Structural Interpretation." *American Sociological Review* 62(3):465–80.

Bonilla-Silva, Eduardo. 2001. *White Supremacy and Racism in the Post-Civil Rights Era*. Lynne Rienner Pub.

Bonilla-Silva, Eduardo. 2017. *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*. Fifth edition. Rowman & Littlefield Publishers.

Bosk, Emily Adlin. 2015. "All Unhappy Families: Standardization and Child Welfare Decision-Making." PhD Thesis. University of Michigan.

Bosk, Emily Adlin. 2018. "What Counts? Quantification, Worker Judgment and Divergence in Child Welfare Decision-Making." *Human Service Organizations: Management, Leadership: Governance* 42(2):205-224.

Brune, Susan. 1994. "Conflict and Conciliation: A Review of the Black Action Movement Strike at the University of Michigan." *Michigan Journal of Political Science* 5:39–66.

Children's Research Center (CRC). 1999. *The Improvement of Child Protective Services with Strucutred Decision Making: The CRC Model.* National Council on Crime & Delinquency.

Children's Research Center (CRC). 2008. *The Structured Decision-Making Model: An Evidenced-Based Approach to Human Services.* National Council on Crime & Delinquency.

Dawes, R. M., D. Faust, and P. E. Meehl. 1989. "Clinical versus Actuarial Judgment." *Science* 243(4899):1668–74.

Drake, Brett, Jennifer M. Jolley, Paul Lanier, John Fluke, Richard P. Barth, and Melissa Jonson-Reid. "Racial bias in child protection? A comparison of competing explanations using national data." *Pediatrics* (2011): peds-201

Duncan, Otis Dudley, Lloyd E. Ohlin, Albert J. Reiss, and Howard R. Stanton. 1953. "Formal Devices for Making Selection Decisions." *American Journal of Sociology* 58(6):573–84.

Espeland, Wendy and Mitchell Stevens. 2008. "A Sociology of Quantification." *European Journal of Sociology* 49(3):401–36.

Font, Sarah. 2013. "Service Referral Patterns Among Black and White Families Involved With Child Protective Services." *Journal of Public Child Welfare* 7(4):370–91.

Font, Sarah., Lawrence Berger, and Kristen Slack. 2012. "Examining racial disproportionality in child protective services case decisions." *Children and youth services review* 34(11): 2188-2200.

Font, Sarah, and Kathryn Maguire-Jack. 2015. "Decision-making in child protective services: Influences at multiple levels of the social ecology." *Child abuse & neglect* 47: 70-82.

Gambrill, Eileen and Aron Shlonsky. 2001. "The Need for Comprehensive Risk Management Systems in Child Welfare." *Children and Youth Services Review* 23(1):79–107.

Golash-Boza, Tanya. 2016. "A Critical and Comprehensive Sociological Theory of Race and Racism." *Sociology of Race and Ethnicity* 2(2):129–41.

Gordon, Leah. 2015. *From Power to Prejudice*. University of Chicago Press.

Grove, William, David Zald, Boyd Lebow, Beth Snitz, and Chad Nelson. 2000. "Clinical versus Mechanical Prediction: A Meta-Analysis." *Psychological Assessment* 12(1):19.

Harcourt, Bernard. 2006. *Against Prediction*. University of Chicago Press.

Harcourt, Bernard. 2015. "Risk as a Proxy for Race: The Dangers of Risk Assessment." *Federal Sentencing Reporter* 27(4):237–243.

Hatry, Harry. 1997. "Outcomes Measurement and Social Services: Public and Private Sector Perspectives." Pp. 3-19 in *Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods*, edited by Edward Mullen and Jennifer Magnabosco. National Association of Social Workers Press.

Hirschman, Daniel, Ellen Berrey, and Fiona Rose-Greenland. 2016. "Dequantifying Diversity: Affirmative Action and Admissions at the University of Michigan." *Theory and Society* 45(3):265–301.

Hirschman, Daniel and Ellen Berrey. 2017. "The Partial Deinstitutionalization of Affirmative Action in U.S. Higher Education, 1988 to 2014." *Sociological Science* 4:449–68.

Hwang, Jackelyn, Michael Hankinson, and Kreg Steven Brown. 2015. "Racial and Spatial Targeting: Segregation and Subprime Lending within and across Metropolitan Areas." *Social Forces* 93(3):1081–1108.

Kahn, Jonathan. 2017. "Pills for Prejudice: Implicit Bias and Technical Fix for Racism." *American Journal of Law & Medicine* 43(2–3):263–78.

Kalev, Alexandra, Frank Dobbin, and Erin Kelly. 2006. "Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies." *American Sociological Review* 71(4):589–617.

Karabel, Jerome. 2005. *The Chosen*. Houghton Mifflin Harcourt.

Killgore, Leslie. 2009. "Merit and Competition in Selective College Admissions." *The Review of Higher Education* 32(4):469–88.

Krase, K. 2013. "Differences in racially disproportionate reporting of child maltreatment across report sources." *Journal of Public Child Welfare 7:* 351–369.

Krippner, Greta R. 2017. "Democracy of Credit: Ownership and the Politics of Credit Access in Late Twentieth-Century America." *American Journal of Sociology* 123(1):1–47.

Lemann, Nicholas. 1999. *The Big Test*. Farrar, Straus and Giroux.

Lewis, Amanda E. and John B. Diamond. 2015. *Despite the Best Intentions*. Oxford University Press.

Lindsey, Duncan. 1992. "Reliability of the Foster Care Placement Decision: A Review." *Research on Social Work Practice* 2(1):65–80.

Meehl, Paul E. 1954 [2015]. *Clinical Versus Statistical Prediction*. Reprint edition. Echo Point Books & Media.

Miller, M. 2011. *Structured Decision-Making risk assessment: Does it reduce racial disproportionality in Washington's child welfare system?* Washington State Institute for Public Policy. Available at http://www.wsipp.wa.gov/ReportFile/1086 .

Moore, Wendy Leo and Joyce Bell. 2011. "Maneuvers of Whiteness: 'Diversity' as a Mechanism of Retrenchment in the Affirmative Action Discourse." *Critical Sociology* 37(5):597–613.

New York Times. 2013. "How Minorities Have Fared in States With Affirmative Action Bans." Available at http://www.nytimes.com/interactive/2013/06/24/us/affirmative-action-bans.html .

Okpych, Nathanael and James L. H. Yu. 2014. "A Historical Analysis of Evidence-Based Practice in Social Work: The Unfinished Journey toward an Empirically Grounded Profession." *Social Service Review* 88(1):3–58.

Osterling, K. L., D'andrade, A., & Austin, M. J. 2008. "Understanding and addressing racial/ethnic disproportionality in the front end of the child welfare system." *Journal of evidence-based social work* 5(1-2): 9-30.

Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market A Field Experiment." *American Sociological Review* 74(5):777–99.

Pelton, Leroy H. 2015. "The continuing role of material factors in child maltreatment and placement." *Child Abuse & Neglect* 41:30-39.

Pettit, Becky and Bruce Western. 2004. "Mass Imprisonment and the Life Course: Race and Class Inequality in US Incarceration." *American Sociological Review* 69(2):151–169.

Poon, Martha. 2009. "From New Deal Institutions to Capital Markets: Commercial Consumer Risk Scores and the Making of Subprime Mortgage Finance." *Accounting, Organizations and Society* 34(5):654–74.

Porter, Theodore M. 1995. *Trust in Numbers*. Princeton University Press.

Putnam-Hornstein, Emily, Barbara Needell, Bryn King, and Michelle Johnson-Motoyama. 2013. 'Racial and ethnic disparities: A population-based examination of risk factors for involvement with child protective services', *Child abuse & neglect*, 37:33-46.

Quillian, Lincoln. 2006. "New Approaches to Understanding Racial Prejudice and Discrimination." *Annual Review of Sociology* 32:299–328.

Ray, Victor. 2019. "A Theory of Racialized Organizations." *American Sociological Review*. Forthcoming.

Reich, Jennifer. 2005. "Fixing families." *Parents, Power, and the Child Welfare System. New York & London: Routledge*.

Reskin, Barbara. 2012. "The Race Discrimination System." *Annual Review of Sociology* 38(1):17–35.

Roberts, Dorothy E. 2003. "Child Welfare and Civil Rights." *University of Illinois Law Review* 171-182.

Rossi, P.H., J. Schuerman, and S. Budde. 1999. "Understanding decisions about child maltreatment." *Evaluation Review*, 23: 579-98.

Rugh, Jacob S. and Douglas S. Massey. 2014. "Segregation in Post-Civil Rights America." *Du Bois Review: Social Science Research on Race* 11(2):205–232.

Russell-Brown, Katheryn. 2018. "The Academic Swoon Over Implicit Racial Bias: Costs, Benefits, and Other Considerations." *Du Bois Review* 15(1):185–93.

Schwalbe, Craig S. 2009. "Risk Assessment Stability: A Revalidation Study of the Arizona Risk/Needs Assessment Instrument." *Research on Social Work Practice* 19(2):205–13.

Sedlak, A. J., and Diane Broadhurst. 1996. "Executive summary of the third national incidence study of child abuse and neglect (NIS-3)." *National Center on Child Abuse and Neglect (DHHS), Washington, DC*.

Shlonsky, Aron and Dennis Wagner. 2005. "The next Step: Integrating Actuarial Risk Assessment and Clinical Judgment into an Evidence-Based Practice Framework in CPS Case Management." *Children and Youth Services Review* 27(4):409–427.

Simon, Jonathan. 1988. "The Ideological Effects of Actuarial Practices." *Law & Society Review* 22(4):771–800.

Starr, Sonja. 2014. "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination." *Stanford Law Review* 66:803.

Stein, Theodore and Tina L. Rzepnicki. 1983. *Decision Making at Child Welfare Intake: A Handbook for Practitioners*. Child Welfare League of America, Inc.

Stone, Deborah. 1993. "The Struggle for the Soul of Health Insurance." *Journal of Health Politics, Policy and Law* 18(2):287–317.

Stulberg, Lisa. and Anthony Chen. 2014. "The Origins of Race-Conscious Affirmative Action in Undergraduate Admissions: A Comparative Analysis of Institutional Change in Higher Education." *Sociology of Education* 87(1):36–52.

Timmermans, Stefan and Steven Epstein. 2010. "A World of Standards but Not a Standard World: Toward a Sociology of Standards and Standardization." *Annual Review of Sociology* 36:69-89.

Williams, Richard, Reynold Nesiba, and Eileen Diaz McConnell. 2005. "The Changing Face of Inequality in Home Mortgage Lending." *Social Problems* 52(2):181–208.

Wooten, Melissa and Lucius Couloute. 2017. "The Production of Racial Inequality within and among Organizations." *Sociology Compass* 11(1):1-10.