

**Redefining Merit Through New Routines: Holistic Admissions Policy Implementation in
Graduate Education**

Julie Posselt¹, Deborah Southern², Theresa Hernandez³, Steve Desir¹,
Fatima Alleyne⁴, and Casey Miller⁵

¹ Rossier School of Education, University of Southern California

² School of Education and Information Studies, University of California, Los Angeles

³ Department of Educational Leadership, Northern Arizona University

⁴ American Physical Society

⁵ College of Science, Rochester Institute of Technology

Manuscript accepted for publication in *Educational Evaluation and Policy Analysis*

Redefining Merit Through New Routines: Holistic Admissions Policy Implementation in Graduate Education

Selective educational organizations from preschools to graduate schools have begun reconsidering inherited admissions policies, including standardized test requirements and practices for reviewing files (Bennett, 2021; Langin, 2019). As long as selective admissions have existed, typical policies have institutionalized racial, class, and gender inequities in enrollment, conditioning subsequent inequities in degree attainment and the labor market (Baker et al., 2018; Bastedo & Jaquette, 2011). Admissions may also reinforce racial stereotypes and racialized conceptions of risk, belonging, and merit (Posselt, 2016; Warikoo, 2018), which matter both as grounds for selection and because they may permeate faculty-student interactions and other educational processes (Canning et al., 2019; Harper, 2010; McGee, 2020). However, as Supreme Court decisions have made salient, public policy constrains the admissions processes that colleges and universities may use to pursue racial diversity and reduce inequalities.

The US Supreme Court's majority opinion in *Grutter vs. Bollinger* (2003) endorsed holistic review as an approach to admissions that is aligned with diversity goals. Holistic review remains lawful under the Court's majority opinions in *Students for Fair Admissions vs. President and Fellows of Harvard College* (2023) and *Students for Fair Admissions vs. University of North Carolina* (2023). Holistic review involves assessment of individuals' "academic ability coupled with a flexible assessment of applicants' talents, experiences, and potential to contribute to the learning of those around them," by taking into account "all of the information available in their file" (539 U.S. 315). It is used in private and charter K-12 admissions, scholarship and fellowship selection, and faculty hiring, as well as postsecondary admissions. Holistic review gained support during the COVID-19 pandemic with surges in test-optional admissions and

assignment of pass-fail grades (Perez et al., 2021). Disciplinary societies such as the American Astronomical Society and American Sociological Association (ASA) have endorsed it, with ASA arguing, “admissions committees [need] to engage in continued self-scrutiny. We also need more research on the mechanics and outcomes of holistic review” (ASA, 2021).

Advocates of holistic review cite the value of 1) looking beyond academic metrics with disparities by race, socioeconomic status, and/or gender (Bastedo et al., 2018); 2) aligning standards of quality with commitments to diversity and equity (Garces, 2014); and 3) assessing socio-emotional qualities that affect professional success and organizational health (Barcelo et al., 2021). Evidence from a carefully managed race-conscious, holistic undergraduate admissions policy at the University of Texas found that it yielded significant socioeconomic diversity among Black and Latinx students, described as “diversity within diversity” (Garces, 2014).

In the eight states that have had affirmative action bans, enrollment by graduate students of color declined (Garces, 2012), with notable losses in natural sciences and engineering (Garces, 2013). Losses appear due to both constraints on the direct consideration of race and how a repressive public policy environment affects professionals’ behavior (Garces et al., 2021). Under race-neutral policy, undergraduate admissions officers report little clarity about how to factor their knowledge of applicant race into evaluations (Poon et al., 2020). To avoid litigation risks, they often err on the side of avoiding discussion of it (Garces & Bilyalov, 2019). This ambiguity may be in part because holistic review does not presume consideration of race or any specific criteria. It is a general approach that institutions must design policies around themselves, then prepare decision-makers to carry out. Whether and how a process to admit more diverse students can do so without considering race may depend upon local designs and embedded routines.

Research Focus

To examine these issues, we explored thirteen PhD programs' implementation of holistic admissions policy from the perspectives of department chairs, graduate admissions chairs, and directors of graduate studies who were responsible for the policy change. Our data collection also included an elicitation exercise that highlights how the COVID-19 pandemic raised some professors' sensitivity to contextual factors in relation to admissions policy reform. We wanted to understand: What criteria and evaluation practices change in the transition to holistic review? How does holistic review stretch or replace established paradigms? What challenges do leaders face in managing change?

Our sample is drawn from 24 natural science and engineering PhD programs in six research universities who were partners of the California Consortium for Inclusive Doctoral Education (C-CIDE), the pilot project for what is now a national Equity in Graduate Education Consortium. In both its pilot and present forms, the Consortium has been aimed at building the capacity (i.e., knowledge, skill, and motivation) of faculty, PhD programs, and graduate schools to select and serve students from historically marginalized backgrounds. Participants received professional development and coaching to raise awareness about inequities in admissions; led their colleagues in discussion and efforts to adopt holistic admissions; and participated in research and project evaluation.

We define equity as reconfiguring structures, cultures, and systems to reduce disparities and empower marginalized groups (Posselt, 2020), and we define equity-minded as the perspective of practitioners who take responsibility for inequities (Bensimon et al., 2016). Like equity, holistic review can mean different things to different people (Bastedo et al., 2018). The Consortium has, over time, refined a model of holistic review that is comprehensive,

contextualized, systematic, and equity-minded (Miller and Posselt, 2020). It was also purposefully developed with institutional partners in California, anticipating that a variant on Proposition 209 would, in time, become the national norm. Our findings show adoption of holistic review involved three common changes to typical routines: delegitimizing GRE scores as a core criterion, developing new review routines via rubrics, and learning new cognitive routines of contextualized assessment. Administrative data from an external evaluation of C-CIDE indicated that PhD programs in the project significantly increased the number and fraction of racially minoritized students who applied, were admitted, and enrolled.

Our findings about the implementation of equity-minded holistic admissions make a critical contribution to the literature and national dialogue at this point in higher education history. With rare exception (Hirschmann et al., 2016), research on holistic review focuses on changes in selectivity or student composition outcomes (Belasco et al., 2015; Bennett, 2021). How holistic admissions is designed and implemented may be driving these outcomes, however, and these processes are therefore critical to analyze as holistic review spreads in undergraduate, graduate, and professional education alike. Our work directly informs this policy dialogue.

Literature Review

Holistic review is new as a focus of research (Bastedo et al., 2018; Rosinger & Ford, 2020; Posselt, et al., 2020), though national associations in medical and graduate education have recommended it for more than a decade (American Association of Medical Colleges (AAMC), 2010, 2014; Council of Graduate Schools, Kent & McCarthy, 2016; NASEM, 2018). Common features in holistic review models to date include reducing the weight of or eliminating test score requirements, incorporating new criteria such as student experiences, and rating applicants individually. AAMC's model advises, "Experiences, attributes, and academic metrics are

evaluated and scored in a systematic and consistent manner across the entire applicant pool, with due consideration to the demonstrated validity of various criteria in predicting success in both medical school and medical practice" (Witzburg & Sondheimer, 2013, p. 1566).

Perhaps owing to AAMC's early proposal of this model, Wilson et al. (2018) conducted research in which three reviewers from an admissions committee scored each applicant file on research experience, recommendation letters, personal statements, and quantitative measures. Reviewers were trained to identify distance traveled (i.e., the academic growth one has achieved from their entry into college, which can be notably different between marginalized and non-marginalized groups) opportunity structures, and obstacles that students may have overcome. Another study found an important role in post-baccalaureate holistic review for individual ratings, but instead of rating elements of the application, they rated aspects of individual applicants such as distance traveled (Barcelo, 2021). To clarify the motives for promoting holistic review, we review research about conventional admissions criteria and processes, and the affordances and limits of holistic review as an alternative approach.

Rethinking Admissions Criteria

Holistic review, test-optional/no-test policies, and race-conscious admissions are often advocated at the graduate level because typical admissions priorities and practices have institutionalized patterns of racial, gender, and socioeconomic exclusion (Attiyeh & Attiyeh, 1997; Miller & Stassun, 2014; Posselt et al., 2019). GRE scores, undergraduate grade point average, and undergraduate institution selectivity were the three strongest predictors of graduate school admission in the largest multidisciplinary study to date (Attiyeh & Attiyeh, 1997). The same factors, plus strength of letters of recommendation, were the strongest predictors of admission in a study of economics PhD programs (Jones et al., 2020); grades and GRE scores

were the strongest predictors of admission in a study of six PhD programs in physics (Posselt et al., 2019).¹ However, GRE scores are unevenly distributed by race and gender (Miller & Stassun, 2014), and access to selective colleges is imbalanced across lines and intersections of race, gender, and socioeconomic status (Baker et al., 2018; Bastedo & Jaquette, 2011; Posselt et al., 2014; Bielby, 2012; Thornhill, 2019). Weighting GRE scores and college prestige as admissions criteria thus privileges groups already overrepresented in doctoral education and the scientific labor force: men, applicants who are white and Asian, and those from middle- and upper-class families.

Reliance on unequally distributed criteria perpetuates inequalities, and misattribution of what the criteria signal can legitimize those inequalities (Bourdieu & Passeron, 1990). Faculty often lean on narratives about risk rooted in faulty assumptions about the relationship of GRE scores and student performance to justify rejecting applicants with low GRE scores (Posselt, 2016). Meta-analyses indicate that scores are often associated with grades in graduate school, but not with longer term outcomes (e.g., Kuncel & Hezlett, 2007). College grades are a stronger predictor of PhD completion than GRE scores (Miller et al., 2019, 2020; Verostek et al., 2021), but are imperfect proxies for the skills and dispositions that enable success in research.

Some propose assessing socio-emotional competencies to complement academic metrics. Time management, achievement orientation, teamwork, and perseverance are examples of skills that many graduate programs already seek in admissions, albeit often unsystematically (Kyllonen, 2005; Kyllonen et al., 2011). First published in a review of research that identified factors associated with student success among students of color (Sedlacek & Brooks, 1976), assessing socio-emotional skills can enhance the validity of selection processes and increase

¹ So important are metrics in physics PhD programs that with only undergraduate GPA and physics GRE score, Young and Caballero (2019) predicted admission with 75% accuracy.

diversity in who is selected (Kyllonen et al., 2005; Ployhart & Holtz, 2008; Sackett et al., 2001). They may predict academic and job performance above and beyond typical measures (Schmidt & Hunter, 1998). These qualities tend to be orthogonal to academic metrics, and most established measures of them show negligible differences by race, gender, or culture of origin (Feingold, 1994; Foldes et al., 2008; Oswald & Hough, 2011).

Rethinking Admissions Decision-Making

Research reviewed thus far implies that a more equitable approach to admissions may involve reducing the weight of (or eliminating) some factors, while adding new ones in a set of criteria that is aligned with diversity goals. However, recent research on college admissions suggests it may not be this simple: Bastedo et al. (2018) identified three different approaches to holistic review in a sample of undergraduate admissions officers: *whole file* considered all parts of the application; *whole person* accounted for a variety of academic and personal qualities; and *whole context* considered information in the context of applicants' opportunities. Only those in the *whole context* condition were more likely to admit students with lower socioeconomic status (Bastedo et al., 2018).² How faculty interpret the information they have about applicants matters.

Whole file and whole person approaches both center on *what* is considered. A panel analysis found that weighting extracurricular involvement (especially forms tied to race and class) and undisciplined assessment of essays, interviews, and letters of recommendation “may reproduce and even exacerbate inequitable enrollment patterns” (Rosinger et al., 2021, p. 48). Considering background factors like first-generation status without reweighting other criteria is also unlikely to shift enrollment on its own but may help in a broader strategy that includes

² The value of attending to one's context of prior opportunities is also indicated in views that assessing one's “distance traveled” provides a gauge to whether an applicant will make the most of resources that a graduate program provides (Ray & Brown, 2015).

recruitment and financial aid (Rosinger et al., 2021). The Bastedo et al. and Rosinger et al. studies suggest a need to change both *what* information is reviewed and *how* review occurs.

Here, it is useful to consider differences between undergraduate and graduate admissions. While the former is highly centralized and formalized, the latter operates as *ad hoc* policy: inherited, decentralized, and informal (Klitgaard, 1985). Judgments can vary from year to year, reviewer to reviewer, and even applicant to applicant. Absent shared evaluation norms or standards, faculty often compare applicants with one another (Posselt, 2016), which is problematic: biases thrive in unstructured conditions (Hsee, 1995; Milkman et al., 2008), and it privileges those who have had the most opportunities (Lamont, 2009; 2012). Faculty often have explicit biases, such as preferences for students who need less mentoring or come with the research training that PhD programs exist to provide (Posselt, 2016). Ethnographic evidence from ten highly ranked PhD programs' admissions processes revealed that racial inequalities are institutionalized not only sorting applicants with unevenly distributed metrics and credentials, but also in coded evaluations of language in letters of recommendation, interviews, and committee deliberations (Posselt, 2016). Approaches to holistic review may thus reproduce inequalities if they do not confront racialized, gendered, or classed ways that we think and talk about what makes a desirable graduate student.

To summarize, research has examined how typical admissions processes undermine opportunities for historically marginalized groups. Experimental evidence from undergraduate admissions indicates that contextualized review can be a lever for more equitable outcomes; however, there is little to no research about how faculty adopt new admissions policies. This question is critical to understanding the diffusion of holistic review and its potential impacts.

Theoretical Framework

Admissions involves individual evaluation and organizational decision-making; it reflects a combination of formal policy, standard practice, and individual interpretation. This multi-dimensionality creates an intriguing empirical context for examining change and suggests that new routines may be foci of change in implementing holistic review. Routines are the space where individual discretion is used to animate policy (Lipsky, 1980), and through routines, racialized scripts in the minds of actors shape the distribution of resources, including time, money, and access (Ray, 2019). In this section we introduce organizational routines and then outline mechanisms through which routines may institutionalize or disrupt inequities. These perspectives provide a useful foundation for examining the implementation of holistic review and its potential and limits for disrupting inequities.

Routines lend structure, rhythm, and stability to organizational life (Powell & DiMaggio, 1991). Feldman (2000) defines routines in the spirit of standard operating procedures: “repeated patterns of behavior that are bound by rules and customs and that do not change very much from one iteration to another” (p. 611). Feldman and Pentland (2003) found that how routines are carried out by real people in specific settings (i.e., routines’ performative aspect) may align with or diverge from the articulation of an idealized routine (i.e., the ostensive aspect). Tracking (mis)alignment is useful because outcomes of a routine may diverge from intent if there is divergence between the performative and ostensive dimensions (Feldman & Pentland, 2003).

Routines as Mechanisms for Institutionalizing Inequities

Routines can operate as micro-foundations of inequities through several mechanisms. There is decades of evidence that hiring and admissions policies include evaluative routines which, when applied over a large volume of applications, may institutionalize socioeconomic

inequalities (Bourdieu, 1984; Karen, 1990; Lamont, 2009). Reviewers working quickly without protocols routinely miscategorize elite social and cultural capital as excellence. “Although somewhat antithetical to a merit-based system,” Lamont writes, valuing cultural capital and other “evanescent qualities... are intrinsic to the process of evaluation in academia” (Lamont, 2009, p. 161). How admissions committee members discuss applicants further entrenches privilege. Through evaluative storytelling (Stevens, 2007), applicants from wealthy backgrounds stand a better chance of admission. Reviewers “glue attributes together into coherent and aesthetically compelling composites” (Stevens, 2007, p. 186), and wealthy students are more likely to have resources to craft an application from which reviewers can create a compelling narrative.

Scholars have also proposed several mechanisms through which routines in education institutionalize racial inequities. For example, racially disparate outcomes may result when an ostensibly race-neutral routine is not enacted in a race-neutral way (Ray, 2019; Ray & Purifoy, 2019). When we started this research, for example, one motivation for US PhD programs eliminating GRE scores requirements was that professors’ routine interpretation and use of scores misaligned with ETS guidance (ETS, 2015; Langin, 2019).

Routine interactions in education that have an evaluative component (e.g., teaching, admissions) may also be laced with racial meanings or biases (Posselt, et al. 2020; Rodgers, 2021). Implicit biases and explicit preferences picked up in our racially hierarchical social system (Hall & Morley, 2019) are predictably baked into moments where educators execute professional, subjective judgments, often with little accountability (Eaton et al., 2019; Quinn, 2020). What Ball (2018) describes as “discretionary spaces” (Ball, 2018) are pervasive in graduate education: Should I take time to reply to this email inquiry from a prospective applicant? Does it matter if they are from an MSI with which I am unfamiliar? How should I

interpret the significance of an applicant's leadership experience with Black Lives Matter? Apart from how applicant race is accounted for or not as a criterion, racial bias in the routines used to make meaning of and act on information that arises in admissions may be consequential for institutionalizing inequities.

Finally, educators may be unaware of racially disparate impacts of evaluation tools and practices. Diamond and Lewis (2019) found school disciplinary routines marked by differential selection (i.e., more frequently identifying Black students as requiring discipline) and differential processing (i.e., more harshly punishing Black students). They also found racialized cultural beliefs uphold these disciplinary routines. Specifically, they found that Black students are more likely to be viewed as a threat and that white students deserve the benefit of the doubt. Such underlying beliefs need questioning for routines to contribute to cultural change.

To summarize, theory suggests routines may institutionalize inequities under several conditions: 1) when an ostensibly race-neutral routine is not performed in a race-neutral way; 2) when routines normalize preferences that privilege groups who are already advantaged, 3) when routine interactions are laced with racial biases, and 4) when educators are unaware their routine preferences and tools have racially disparate impacts. However, routines can also provide flexibility (Feldman, 2000; Feldman & Pentland, 2003), which is a source of equity-minded institutional change (Hirschmann et al., 2016). We turn now to summarize these possibilities.

Routines as Mechanisms for Disrupting Institutionalized Inequities

Institutionalized inequalities may be disrupted by changing evaluation and decision-making routines that allocate opportunities and resources. "Meso-level racial change may arise endogenously from changes in organizational routines," Ray (2019) writes, because "individual racial attitudes and discrimination are enabled or constrained by organizational routines" (p. 30).

New routines can operate as tools for institutionalizing new mindsets and behaviors; they systematically, repeatedly change how attention is focused, the information to which we attend, and the meaning we make of information (Cheryan & Markus, 2020). Individual agency in routines allows for innovation, and “if enough people or even a few people who are powerful enough act in innovative ways, their action may have the consequence of transforming structures” (Cheryan & Markus, 2020, p. 4). In short, discretionary spaces can be used for good.

Finally, adoption of tools that routinize shared expectations for evaluation can (re)structure the rules of admissions. Standard operating procedures are produced via the “articulation, codification, and standardisation” of expectations (D’Adderio, 2008, p. 781). Equitable practice can effectively become routine with equity-minded expectations for acceptable practice. Hirschman and Bosk (2020) propose that protocols like admissions rubrics can be designed to address structural inequalities in decision making. They present the University of Michigan’s point system as a tool that advanced racial diversity because its designers intentionally addressed racial disproportionality in various admissions criteria. In a process designed to be holistic, which balances collective activity and individual judgment, routines that disrupt institutionalized inequities may include changing how attention is focused, what people focus on, and new tools for decision making.

Recognizing that routines could describe what organizational actors are substantively changing in the implementation of holistic review, and that new routines would come with both opportunities and limits for racial equity, we posed the following research questions:

1. What new routines are used in implementing holistic review? What do faculty perceive as the affordances and limits of new admissions routines?

2. What barriers emerge in the uptake of more equitable admissions practices? How do leaders manage these barriers?

We framed these questions to uncover organizational and individual level realities and to attend to both what was changing and how leaders experienced it. Analysis in the discussion section explores implications for racial equity that follow from the findings and conceptual framework.

Methods

The context for this research is a research-practice partnership, which the W.T. Grant Foundation (2021) defines as “long-term, mutually beneficial collaborations that promote the production and use of research.” The Consortium pilot included six research universities’ graduate schools/ divisions (University of California Berkeley, Davis, Irvine, San Diego, Santa Barbara and University of Southern California) and 24 STEM PhD programs within them. The programs were open to experimenting with existing policies and practices over three years and to having their processes and outcomes studied. The goals were to facilitate learning and support for aligning admissions policy and practice with values of racial equity and diversity.

Project leaders facilitated two, two-hour workshops about current research on admissions with each program in 2018-2019, then maintained contact with program leaders. This contact included 1) annual surveys of admissions and recruitment practices by an external evaluator (See Table 1), 2) program-level administrative data provided to the external evaluator about applications, admissions, and enrollments (disaggregated by race/ethnicity and gender) prior to and following the workshops, 3) regular reflective memos, 4) one-hour research interviews in Spring 2020 (occurring after the end of the admission cycle during which they attended workshops and began implementing changes), 5) light, email-based coaching on specific issues of admissions, recruitment, and change management, 6) asynchronous resources, and 7)

quarterly, online discussions of relevant social science research (e.g., responding to COVID, using data for equity). We purposefully created multiple means of engagement so partners could choose based on their needs and interests. This paper reports the in-depth qualitative interviews. In-depth interviews that take place through longer-term engagement with participants in their contexts allow researchers to contextualize participant experiences and connect them to relevant structures (Weiss, 1994; Ortiz, 2003). Selected evaluation data are provided as context to describe the sample and set the stage for the findings.

Sample

Sampling occurred at the program and individual levels. 24 PhD programs joined the RPP in part because they identified as “change-ready”; that is, they already recognized an equity problem, agreed about the need for change, and were seeking guidance and resources. Programs’ openness to change offered an important condition for studying holistic review implementation. We used purposive, criterion sampling (Creswell, 2009; Maxwell, 2013) to invite each of the program’s faculty leaders to participate in a research interview centered on their experiences implementing program changes. All were faculty and either the current chair of admissions or director of graduate studies, depending on who was most involved with leading changes to admissions. Sixteen representatives from 13 programs across five partner universities agreed to participate; most who declined cited increased time demands due to COVID-19.

The sample came from five California universities, four public and one private, all R1: Doctoral Universities with very high research activity. Of the 13 programs represented in the interviews, seven are from biological sciences, one is from engineering, and two are from physical sciences. Table 1 displays additional data about admissions practices in the full set of 24 PhD programs that were part of the RPP during data collection for this paper. Appendix 1

compares the number and fields of participating and non-participating programs. Biological science programs are overrepresented in the research sample relative to the programs in the partnership overall, while engineering programs are underrepresented.

(Insert Table 1)

Data Collection

Our data collection focused on individuals' experience leading change and working to mitigate disparate impacts of admissions processes. Data collection consisted of one-hour, semi-structured video interviews in Spring 2020. Participants and researchers conducted interviews from their homes due to COVID-19 work-from-home requirements. For each interview, two interviewers were present, with one designated as the leader and the other responsible for taking notes and managing technical aspects of the interview. Two of the three researchers involved in data collection were not involved in facilitating workshops. Our protocol included questions about programs' graduate admission processes and efforts to change them, including policies, organizational structures, and difficult colleagues. We also asked about their leadership experience and the contexts for change, as well as inquiring about what routines changed in implementing holistic review. By asking open-ended questions with follow-up probes, we elicited detailed, reflective, and narrative data (Weiss, 1994; Ortiz, 2003).

To create a space for processing the unfolding COVID-19 pandemic, the interview protocol also included a transcript review elicitation activity. Elicitation techniques draw out insights about tacit and procedural knowledge that participants might be reluctant to share or have limited experience discussing (Barton, 2015; Johnson & Weller, 2002). In our case, we knew that COVID-19 would likely present an exogenous shock to traditional application review; the transcript review provided a window into the pandemic's implications for norms about the

importance of college grades and what grades were desirable (Maitlis & Sonenshein, 2010). During this portion of the interview, each participant was provided a real transcript that had been stripped of personal information, and in which the final semester's marks (Spring 2020) were changed to reflect pass/fail grading. We altered three undergraduate transcripts— one each from biological sciences, physical sciences, and engineering—to align with the interview participant's general STEM field. We asked faculty to talk us through their routines for reviewing transcripts, including what information they looked for, what caught their attention, and how they would use the student grades to assess admissibility. We probed how they made sense of grades, generally, and pass/fail grades received during COVID-19, specifically. Appendix B provides the prompt and interview questions for this activity.

Data Analysis

Our analysis followed the constant comparative method (Charmaz, 2014), as adapted from grounded theory (Strauss & Corbin, 1994). Three researchers who interviewed participants separately memoed after each interview to record insights that emerged, toward development of sensitizing concepts (e.g., ambiguity of strong applicant qualities, tension around the GRE, red flags on transcripts). After the interviews were professionally transcribed and checked by the research team, two of the researchers utilized sensitizing concepts and open coding in Dedoose for round one of analyzing all transcripts (Saldaña, 2013). Additional codes created included contextualization, challenges, and admissions policy. The two lead analysts practiced intercoder reliability to check for shared understanding of the data and alignment of codes (Saldaña, 2013).

The full research team then discussed findings from the first round of analysis before the two lead data analysts began axial coding to 1) disaggregate umbrella codes in which there were meaningful distinctions and 2) relate codes to one another (Saldaña, 2013). They again engaged

in intercoder reliability to check for understanding and discuss how codes were distributed across participants (Saldaña, 2013). Here we identified three different dimensions to debates about the GRE's legitimacy, for example, and noticed contextualization not only as a general pattern in the data, but as a routine for interpreting information within the transcript and across the application.

Transitioning to selective coding, the research team discussed axial codes in relation to larger themes, such as challenges with contextualizing applicant qualities. Then, we connected those to our understandings about the role of routines in organizational change. We cataloged mechanisms by which routines might disrupt institutionalized inequalities, especially but not exclusively by race, and then assessed the alignment of our findings with those mechanisms. We summarized early findings with data at this point, providing relevant quotes from participants as evidence. After completing analysis, the research team also populated reports in Dedoose to check interpretations of the data, reviewing code overlap and frequency across participants.

Finally, as an equity check, one analyst re-reviewed all transcripts for mentions of race, racial/ethnic groups, diversity, (in)equity, and inclusion. This surfaced a pattern of race-neutrality in responses to open-ended questions about admissions and recruitment practices. That is, only a few people used race-conscious language or expressed racial awareness. As our findings discuss, even when holistic review improves compositional diversity, a limitation to holistic review where policy precludes explicit consideration of applicant race may be to suppress discussion of race necessary to correct stereotypes and racialized conceptions of merit.

Positionalities

Our research team represents diverse disciplines, professional roles, and social identities, which informed our interpretation of data and engagement with the research. Four have research expertise in higher education, while a physicist and an engineer are familiar with the work of

rethinking admissions in STEM. We pulled from diverse experiences in graduate education as faculty, a director, associate deans, research assistants, and doctoral candidates, which provided us with a broad perspective about the dynamics of attempting change. Our social identities offered similar breadth of perspective: the team includes women, men, and non-binary individuals who identify as Black, Latinx, and white. Shared leadership, flexibility, and mutual support have been crucial as COVID-19 and a rise in police brutality introduced stressors for each of us. We consider these more than incidental to our work; in a research- practice partnership dedicated to equity, our team strives to embody principles of equity and inclusion that we hope to see normalized in the STEM community.

Trustworthiness

The research team adjusted the pace and division of labor of our activities throughout the COVID-19 pandemic to ensure mutual support and rigor in the research. To ensure trustworthiness, we practiced intercoder reliability in all three rounds of analysis, as described above (Creswell, 2009; Maxwell, 2013). We met regularly to discuss findings, interpretations, and what we were learning from the data. Extended engagement with programs via the RPP enabled contextualization of our cross-sectional interview data and a deep understanding of their change processes. Finally, we practiced member checking by discussing our preliminary findings with selected research participants and the campus liaisons (Creswell, 2009; Maxwell, 2013).

Limitations

Limitations include data collection during the COVID-19 pandemic. It presented a unique opportunity to explore how the pandemic affected interpretations of student records, but participation was lower than desired due to the impact of the pandemic on peoples' lives. There could be selection bias favoring faculty with time, resources, and good health in Spring 2020, but

the majority of programs in the Consortium were part of our sample. One member of the research team was involved in both workshops and half of the interviews; it is possible that their presence in interviews might have influenced participant comments. Relying upon interview data to document experiences with holistic review implementation may also be a limitation, and while our interview protocol asked select questions about race, it was not a focus. Under Implications for Research, we recommend future research to complement this work.

Findings: New Routines in Holistic Review

The transition to holistic review in these 13 programs involved delegitimizing previously unquestioned evaluation routines and protecting the possibility of new routines by managing the dialogue around change. The most common changes involved altering their evaluation criteria, adopting evaluation rubrics, and developing new habits of mind for interpreting data across elements of the application. We found two areas where holistic review especially stretched the current paradigms of these STEM scholars: First, learning to read academic preparation metrics in the context of social trends required confronting deeply held beliefs about the power and precision of numbers. Also challenging was developing shared standards when individuals had historically been left to judge files as they saw fit. To navigate these challenges, our participants developed new tactics in change management; they also deferred to scientific logics of data-driven decision making, as well as imperatives of access, diversity, equity, and inclusion. This section reviews patterns in new routines, associated challenges, and how leaders navigated them (See Table 2). The discussion will revisit the findings in light of the conceptual framework's mechanisms for interrupting institutionalized inequalities via evaluative routines.

(Insert Table 2)

Revisiting the GRE: Structural Change via New Application Routines

Every program revisited their policy on whether to require GRE scores for admission. Like any policy change, the challenges in revisiting the GRE were often layered: programs needed support from faculty; bylaws might need amendment; and in some cases, permission to deviate from institutional policy might be needed from a graduate school or faculty senate. The chair of an earth sciences program shared, "You talk to people about the GREs and if [discussion about changing the requirement] dies in some committee or in some office above you, you have to decide whether that's the hill you want to die on versus something else." Leading a proposal to change GRE policy came with distinctive difficulties to "deal with," as many in the sample put it: bureaucratic obstacles; spending political capital; or heated discussion with colleagues. Still, our external evaluation showed 82% of participating programs eliminated GRE requirements, with the greatest proportion attributing this change to learning via C-CIDE, in addition to peers' behavior in their field and the COVID-19 pandemic. One engineer explained:

We made a big push to do more holistic application reviews, I think in part because of some of the workshops that you've been involved in and some of the materials we had seen. We had for a long time, I think, over-emphasized GRE, and now we essentially ignore it, unless it's devastatingly bad.

Participants' descriptions of deliberations about the GRE revealed three common questions about the legitimacy of the exam as a criterion: Some discussed the GRE as a *data point* in representing desired student qualities; many discussed the GRE as creating *barriers* for applicants; and some discussed how *peer programs'* changing GRE policies affected their own decision. As we discuss, faculty leading discussions to revise GRE requirements grew adept in managing debates about the GRE's legitimacy on these dimensions. Such discussions were necessary to surface and correct misperceptions and negotiate outcomes with broad support.

GRE Scores as a Data Point

A critical aspect of discussions about the GRE's legitimacy concerned whether scores help predict student outcomes or represent applicants' academic abilities. Data, of course, is a foundation of scientific inquiry, and many scientific communities aspire to bring to PhD admissions the same data-driven, dispassionate, and objectivist disciplinary logic of research in the natural sciences (Posselt, 2015). One participant described a colleague's insistence that the GRE requirements be continued: "They said, 'Well, the GRE score is another piece of data. And why, as scientists, would we not collect data in making this decision? We should be collecting as much information as we have access to.'" By deferring to their shared identity "as scientists," this colleague tried to make the case that they "should be collecting" the GRE scores as "another piece of data." An earth scientist in another department shared: "We have some faculty who were furious that we were losing that piece of data. And I'm like, 'Well, yeah, you're losing a piece of data, but it's flawed data. Are you really – do you want to really include that in your data set?'" Here we notice their observation that, as in research, not all data in admissions is of equal worth.

In debating scores as data points, faculty often turned to research. A biologist said,

Over the past few years, there have been a lot of studies looking at whether or not GRE scores correlate with success in graduate school. And they, by and large, show that they don't, that GRE scores are not a predictor of success in graduate school.

Participants expressed concern that continuing to require or weigh the GRE could have negative consequences for the diversity of the short-list. For colleagues who wanted to continue requiring scores, a common argument was their perceived utility for gauging an applicant's quantitative skills. A biologist shared, "The quantitative skill of the student is important in our field ... we do need a student to have basic quantitative skills such as how to calculate a composition of a solution." Other faculty used GRE scores to judge applicants from unfamiliar institutions.

To evaluate the GRE as a data point, eight programs took it upon themselves to conduct their own program-level analyses. Five programs tested their need for the GRE through a trial in which scores were required of applicants but were either not visible to reviewers until the final round of deliberations, or not at all. This gradual approach helped make clear to faculty how their decisions would look without scores. “As we all suspected,” one put it, “the GRE really didn't play a big role in determining who was going to get the interview.” Two programs conducting their own predictive validity analyses eliminated score requirements. An advocate for eliminating GRE scores in an engineering program explained that they “correlate[d] about six or seven years' worth of students' GRE scores to their grades in the first year graduate courses that we teach, and there was zero correlation. I mean, it was negative in some cases, and it also didn't predict outcomes.” Here and in other participating programs, the availability of local data that aligned with trends in the research literature helped facilitate a policy change.

Barriers Created by the GRE

A second axis in deliberations about the GRE concerned barriers to equity and inclusion created by using scores. This included costs of taking the exam, sending scores to prospective programs, and taking test-preparation courses. One biological sciences program decided the exam's financial burden alone warranted its removal as an admissions requirement:

The conclusion that we came to is that it prohibits a lot of the kind of students that we should be serving from even applying in the first place. Because if you have very limited financial resources, you can scrape together enough to take the test, and then you get to send it to three places, right?... There might be people that are really, really good students that weren't even sending their GREs to us, or weren't applying because they couldn't afford to send them to multiple places.

Faculty also expressed concerns about consequences for diversity. A professor in the biological sciences noted, “If anything, GRE scores are a mechanism for excluding students that maybe have had less traditional academic backgrounds or fewer financial resources to take test prep

classes.” In an earth science program, rethinking the GRE “was an issue driven first by our desire to be more diverse... [We] realized how important the GRE is in terms of being a biased sort of gatekeeper.” An earth scientist in another department that still required the GRE described “two camps” pushing for change: One camp pointed to scores’ weak predictive validity, summed up as “it’s not as if the people we’re admitting have such amazing GREs.” A second camp centered on barriers, summarized as: “Why make students pay for this? It’s biased against certain groups *and* it’s not useful.” Programs also watched and weighed the moves of their disciplinary peers.

Peer Program Policy

The national movement away from the GRE and standardized testing was an important context for programs’ deliberations. Programs in the biological sciences, where more than half of PhD programs nationally had eliminated requirements (Langin, 2019), were especially inclined to benchmark their GRE policies to those of peers. In at least two programs, fear of fewer applicants to their program was the determining factor in deciding to no longer require the exam:

What drove the unanimous decision to drop the GRE in our department was pressure from other campuses and other programs... In the end, it was peer pressure. So, for the few people who were really against dropping the GRE, they just — they went along for it because they realized, “Oh, okay, well, our applications will go down if we don’t drop it.”

Changes in peer programs could sway faculty who had been opposed to eliminating GRE score requirements. The language of “competitive disadvantage” was used to make the case:

The part of the argument that was not necessarily able to persuade the people that were in favor of keeping the GRE, but maybe resonated with them, is that many PhD programs that we compete with for students had decided to get rid of the GRE. So were we putting ourselves at a competitive disadvantage with those programs by continuing to require it? And so this was an argument that was maybe a little bit more persuasive for the people that didn’t — that didn’t see the GRE as a fundamentally biased piece of information.

Competing with peer programs, and implicitly competing for students of color, motivated faculty to align their policy with their peers’.

Managing Concerns and Challenges: GRE Policy

Managing the debate about the GRE's legitimacy was integral to managing resistance to proposed or actual policy changes. Even with majority support or a completed vote for eliminating scores, debate could continue. The two most common strategies that admissions leaders offered for managing debates and resistance were presenting data about the GRE and leveraging colleagues' support. Some leveraged data collected under new admissions regimes to legitimize changes and draw attention to race and gender. An engineering professor shared his experience trying to cement program reforms by discussing data with a skeptical colleague,

Now that we've done a lot of this quantitative analysis to show what actually predicts success in grad school ... [my colleague] was very surprised when he saw the data; he said, 'Whoa, that really surprised me; I had no idea.' And so then I think he realized he had to trust the reality of it, the data, and just came on board.

A junior biology professor questioned a senior committee member who only wanted to admit students with the highest GRE scores, GPAs, and ample research experience, which this participant knew would privilege “white males.” They offered the committee “success stories” of students who did not have high scores, asking whether it was possible that the committee might have overlooked “this whole group of potentially really good graduate students.” Though as a pre-tenured faculty member they were careful not to directly challenge this senior colleague, they also understood that, “you have to call it out.” By “it,” he meant: “he wasn't being blatantly sexist or racist but by their criteria there was no way anyone other than a white male was going to come into our program, even though fifty percent of the applicants were women.” Data — historical and current, statistical and narrative — are used to negotiate GRE policy change and to surface how admissions criteria can create disparate impact that engrains inequalities.

Leaders also leveraged the support of colleagues in managing resistance to changing GRE policy. In response to a false charge in one biology program that the admissions chair had

unilaterally changed GRE requirements, the chair obtained statements from the full admissions committee expressing their support for the policy change: “They sent their response. And this reduced the barrier for the other people on the admissions committee to express their view... Eventually everybody had sent an email saying what they thought. That ended it.” Though unfounded charges stung the chair, vocal support from fellow committee members allowed adoption of a test-optional, holistic admissions policy to proceed. There is power in numbers.

Adoption and Use of Rubrics: New Routines of Evaluative Practice

GRE policy was rarely the only focus of change. Adoption and use of rubrics was widespread, with participants describing the rubric as a tool that routinized holistic review and consistent consideration of new criteria. Two patterns emerged regarding the role rubrics that served in change management. As a *developmental tool*, some discussed rubrics as a way to apply and incorporate what they were learning from workshops, discussion, and research about holistic admissions into their processes. Biologists from two universities specifically noted the workshops as influences on their adoption of rubrics:

I attended some of the workshops from the grad school and then we developed a rubric, and we tried to follow that rubric at that time. The idea was good. I think it started moving things on, like a more equitable review or more consistent review.

Another biologist’s remarks point to review becoming “equitable” and “more consistent” under rubrics by becoming “more fine-grained:”

The big change for us in the past few years that we've learned from our self-assessment but also from participating in workshops, is we just didn't appreciate the degree of the implicit bias in letters. We also didn't appreciate or know to think about, ‘Wow, this student may have actually had a lot of help on their statement and this student may not have.’ [review became] more fine-grained in how we were looking at the applicant materials, and having the rubric helped a lot. On the rubric, it actually had a place to think about ‘what opportunities did the student have’ and ‘did they take full advantage?’

For those already using rubrics for teaching, their familiarity eased introduction of new criteria such as social-emotional competencies, diversity contributions, and distance traveled (i.e., taking advantage of one's opportunities). Rubrics routinized attention to information that might be overlooked, such as statement of purpose and GPA within major. They also helped reviewers employ common standards instead of comparing applicants against one another.

Admissions leaders also discussed the rubric as a *political tool*. It provided a means of aligning gatekeeping priorities with, as one biologist put it, their "belief system about what grad school should be and the types of people who fit in there." In one case where the graduate school required a rubric, an admissions chair removed the GRE from it as an executive action to prevent misuse of scores. The rubric made change less politically contentious by offering a concrete place where people could see their own priorities represented. In short, in addition to routinizing consideration of new criteria (i.e., the *what* of review), rubrics helped programs navigate politicized aspects of new evaluative routines related to the *how* of file review.

Managing Challenges and Concerns: Rubrics

Rubrics arguably represent a cultural shift in graduate admissions: they bring systemization and shared priorities to an historically *ad hoc* process left to individual judgment. The concerns and challenges that leaders negotiated in moving to rubric-based review involved logistics, quality of available information, and quantification. Though these challenges did not impede the adoption of rubrics, they were part of the organizational learning curve associated with a new routine and were resolved with discussion, data, and/or institutional support.

Logistical Challenges. The most frequently mentioned logistical challenge was increased time required per file review, relative to a system in which whole batches of files could be discarded based on one or two metrics. Participants described these as issues of "manpower" and

“labor”, clear indications that admissions review is experienced as work. These concerns were most salient in programs with large numbers of applications. A related challenge in one program involved carrying out this increased work on the same timeline, which they saw as crucial to preserving the competitiveness of their admissions offers. They resolved the dilemma by moving up the date for distributing files and increasing the number of people involved in review.

Concerns About the Quality & Consistency of Information. Concerns about the quality of rubric-based review revolved around the quality of information about applicants and rubrics’ implicit quantification. Judging how students narrated their diversity contributions and socio-emotional competencies proved particularly difficult. “It’s kind of vague,” as one biologist put it. In personal statements, “students are poets... They don’t go directly to the point.” A common concern in assessing socio-emotional competencies was inconsistency across files in the “information available to properly rate” candidates. In the admissions cycle after a rubric had been adopted, several programs worked with their graduate schools to change application instructions and/or guidance posted on program websites. They incorporated new prompts for the statement of purpose, for example, to elicit student experiences with diversity more consistently and with socio-emotional competencies on the rubric. Such changes began by discussing the origins of their judgments. For example, at least one program had open discussion about differences in how social identities conditioned their interpretations:

I had one person on the committee who is an underrepresented minority and comes from a background where diversity and equity was something she felt strongly about. So she scored people very hard, from a point of view, because she was like, ‘Just writing a few words on a piece of paper doesn’t show me that you’re really committed to this.’ Where other members of the committee would say well, ‘from my point of view... as a white male who has never really dealt with these issues as much, seeing some of these statements on paper and someone who’s engaged in something. I have a different bar.’

Rubric-based review thus induced a sharper recognition of how identity and experience may frame interpretation.

Concerns about Rubric-based Quantification. Our data also raised opportunities and challenges afforded by rubric-based judgment — and its implicit quantification — relative to reliance on externally provided metrics. That an individual’s overall admissibility based on several dimensions of quality could potentially be tallied up made holistic review more accessible and legitimate to participants as scientists. Rubric-based quantification could also enable a rough ranking of applicants to guide subsequent discussion, though not all programs used rubrics in this way. Implicit quantification also introduced new ambiguities and potential for disagreement. There were worries in one program that the rubric might not sufficiently differentiate among applicants, either because of how they had operationalized different levels of a criterion or because they had under-trained the committee. “People didn’t know how to score it so they tended to give everyone the same score,” as one put it.

Another concern raised was how to set weights on various criteria. To name the criteria, operationalize them, and decide whether and how to apply weights involved layers of negotiation and recognition that they were “biased” to privilege some qualities more than others. “We have ten different criteria and some coefficient in front of every one of those,” a biologist explained. “It’s the magnitude, the relative weighting, that reveals our biases.” Including and weighting criteria also challenged faculty to go further and take action that may or may not align to their stated values and preferences. A professor of earth sciences illustrated this:

There was a big argument on the weighting of diversity, equity, and inclusion. There were some faculty who...would kinda give you all the counters against it and then follow the next sentence by saying how great and supportive they were of these types of things, but like, not wanting to actually do anything.

The discomfort of creating and using a rubric for the first time was in large part because it forced colleagues to confront previously unstated preferences.

Most appreciated implicit quantification, but we have disconfirming evidence from a physics program in this study. It initially resisted a rubric because of their epistemological preference for quantitative metrics and view that rubrics did not adequately quantify merit. “I think there are a lot of people who are allergic to the term rubric,” the admissions chair shared ruefully. They positioned this “allergy” in relation to physics’ disciplinary culture:

Physicists are used to the idea that everything has a definite numerical answer and there's a right answer and there's a wrong answer. And I believe that people's reaction to the idea of a rubric surfaces just how subjective these judgments really are. And however much you tell people, ‘Yes, they're subjective and this is the structure of subjective judgment,’ they're not hearing it. They're just hearing a subjective part... It's tied to this issue of quality. They believe there's an objective standard of quality and the physics GRE happens to be a simple way to implement that from some people's point of view. And so that's what I think makes them nervous. We have this discussion in faculty hiring. We have this discussion in all sorts of places.

Their belief that physicists hold “an objective standard of quality” that rubrics do not capture speaks to the cultural shift that may be required for admissions, tenure, or any selection process to be carried out without heavily relying upon metrics. As Lamont (2009) wrote, disciplines vary widely in their comfort with the notion that “excellence” is in the eye of the beholder. However, with further discussion and the move to campus-wide admissions software, which encourages more structured, comprehensive review, this program also shifted to using an evaluation rubric. In a process where evaluation has historically been unstructured, rubrics served multiple purposes: a guide to faculty thinking, a place to apply learning about problems of traditional admissions, and a political tool for ensuring people could see their interests represented.

Contextualizing Information: Routines as New Habits of Mind

Adopting the organizational routine of rubric use shaped the sub-routines that individuals used to interpret information, discern student qualities, and assess admissibility. Consistent with an equity-minded approach to evaluation, they learned to contextualize or, as one put it, “look beneath the surface of information” in a file to consider the statistical error of academic metrics and structural correlates of applicant characteristics. Participants cited project workshops as a site for learning contextualization, due to research presented about the standard error for GRE scores, how score distributions vary by race and gender, and variation in grade point averages between public and private universities. As we describe, contextualization destabilized scientists’ belief in academic metrics as objective or standardized, called into question metrics’ assumed superiority to other information in the file, and compelled appreciation for narratives (e.g., letters and personal statements) to obtain fuller pictures of applicants. Faculty also contextualized information in light of COVID-19, within the transcript and across elements of the application.

Contextualization of Grades during COVID-19

As our interviews began in Spring 2020, campuses were moving *en masse* to online instruction and offering students pass/fail grading options. The transcript review activity we added to our protocol explored how the pandemic affected their judgment of grades. Zero of 14 mentioned racially disparate impacts of the pandemic, but nine faculty expressed they would not think twice about accepting “passes” in lieu of letter grades earned in Spring 2020 due to COVID-19. A biologist said, “The general comment that I’ve heard most is that when we see transcripts and everyone has a bunch of passes in the spring of 2020, everyone knows what happened so it’s going to be a moot point...everyone is in the same boat.” They understood the pandemic as an unprecedented and challenging time, and that pass/fail grades were to be

expected. The other five faculty indicated openness to pass/fail grades during COVID-19, but wanted to understand Spring 2020 in a broader story of the applicant. They interpreted pandemic pass/fail grades by triangulating them with letter grades in previous terms. One explained,

It would definitely change my reading, but then I'd be looking for something...in the personal statement and recommendation letters not to compensate, but give me some context of how the university and student, how the department responded in Spring 2020.

COVID-19 influences this participant's assessment of pass/fail grades but they wanted details about the organizational and student responses to construct a narrative.

Contextualization Within the Transcript

The search for a broader narrative is consistent with faculty looking within the transcript and across the file for information about applicants' academic performance. 70% of professors in the transcript review shared that their assessment of pass/fail grades in a *typical* year would be conditional on the availability of additional information in the transcript and application. One explained, "The first thing is to look at trends. Did they start out slow and then increase over time, or did they have kind of a steady experience through their undergraduate years?"

Participants vocalized questions such as: How has the student typically performed? How did the student perform in upper-division courses? In major courses? Is there a pattern in letter grades? As if solving a puzzle, piecing together answers to these questions presented a fuller picture of pass/fail. One participant explained, "If their transcript showed that they were getting mediocre grades all along and then — and then an F and a bunch of pass/no pass, then I would say, Well, you were actually a mediocre student all along." Such comments suggest a mindset that pass/fail is an anomaly, in which additional information is needed before judgment can be rendered.

Contextualization Across the Application

Participants looked to other application materials to contextualize potentially concerning information. For example, they sought within personal statements extenuating circumstances that explained low grades in key courses, a poor semester, pass/fail grades, or withdrawals. One participant shared that it would help alleviate worry about pass/fail grades during COVID-19 “if in the statement there is something, some explanation.” But upon probing what explanation they would seek, they did not speak specifically of the pandemic:

...maybe the student had a particularly tough time their last semester or wanted to do a more intense research project so they tried to take the pressure off by doing their courses pass/fail. There might be information outside of the transcript that explains the way that the transcript looks.

This participant’s imagining a one-off semester or “more intense research” as explanations that would alleviate their concerns, as opposed to anything related to COVID-19, may exemplify “evaluative storytelling” that unwittingly supports privileged applicants (Stevens, 2007).

Four faculty described contextualization in deficit-oriented ways for applicants whom they imagined as racially minoritized. They sought to explain red flags in the transcript through material in diversity statements or via a trajectory of grades because, “many students, especially diversity students, you know, are adjusting to a new life and meeting the challenges of school and sometimes they don't do well in the beginning of their career,” as one earth science faculty mentioned. How faculty practiced contextualization — where they sought explanations for red flags, the identities they assumed of applicants, and whether data alleviated their concerns—underscores the need for equity-minded training in the interpretive subroutines of holistic review.

Discussion

Our research found that implementation of holistic admissions involved reassessing and changing both what criteria were assessed and how evaluation occurred. It included

delegitimizing previously unquestioned criteria and routines and encouraging new, potentially more equitable, routines through effective change management. In our sample of faculty leaders from STEM PhD programs, implementing holistic admissions involved deliberation and new routines at the levels of organizational policy (i.e., GRE requirements) and practice (i.e., adopting rubrics), as well as individual cognition (i.e., contextualized review).

We found programs striving to disrupt institutionalized inequalities by redefining normative definitions of merit and reconfiguring routines for reviewing files. Initiated as a policy change, holistic review was experienced as a cultural shift away from decentralized, *ad hoc* review anchored in select metrics, and toward systematic, contextualized assessment. More than mere reform, programs saw themselves shifting evaluation paradigms. Some came to recognize social identities as origins for their judgments. Central to implementing holistic review was navigating the politics of this cultural shift and the perceived legitimacy of specific criteria. In doing so, leaders found themselves shaping priorities, shifting discourse, and managing skeptical colleagues. They spoke about cultivating the savviness to manage colleagues' concerns or outright resistance. Policy implementation was thus entangled with change management (Bal et al., 2018; Spillane et al., 2018).

Can Holistic Review Interrupt Institutionalized Inequalities?

What do these findings mean for racial equity? Table 3 connects our findings about holistic admissions to mechanisms discussed in the conceptual framework about how evaluative routines can disrupt inequities. Before we discuss specific connections, there are four patterns to notice in this table. First, of the seven mechanisms we identified in the literature, which are listed in Column 1, our findings included data about all of them except becoming aware of racialized meanings of evaluation criteria (Row 3). Second, it is notable that every one of these

mechanisms could be applied in legal contexts that prohibit race-conscious admissions; this is crucial given that the Supreme Court's recent decisions have effectively made such a prohibition the law of the land. Third, most of the mechanisms concern *awareness, attention, and alignment*, which is consistent with work on equity-mindedness as a foundation for professional practice that disrupts inequities (Bensimon et al., 2016). Finally, there are patterns in Column 2, which we discuss next. (Insert Table 3)

The Potential of Rubrics in Graduate Admissions

A common step in implementing holistic review was developing, adopting, and/or refining an evaluation rubric. That rubrics are mentioned in four different rows of Table 3 suggests multiple ways that rubrics may align with equity goals. First, as presented in Row 6, we found that rubrics enabled programs to direct attention to a broader profile of skills and characteristics in their assessments of merit. Diversity contributions and socio-emotional competencies had long been assessed informally, but adding them to a rubric helped standardize assessment across applicants. These findings align with those of Barcelo et al. (2021), Wilson et al. (2018), and Young and Caballero (2021) who found positive outcomes for equity in changing *what* reviewers pay attention to when they read files.

Row 5 in Table 3 acknowledges that changing routines for *how* attention is focused may also advance equity (Quinn, 2020; Eaton et al., 2019; Hirschman & Bosk, 2020). Participants described rubrics as making review more equitable, consistent, and fine-grained. They also discussed the development of cognitive routines involving contextualization. This involved learning to triangulate information across different parts of the application and to think of groups of data points as indicative of deeper qualities than what any single data point could convey. Professors also learned to regard unequal access to opportunities as a structural context for the

achievements that an application did or did not include. Here, the COVID-19 pandemic enabled them to see in real time how challenges outside of an applicant's control might manifest in a record that falls short of their "ideal."

As D'Adderio (2008) found, we found the creation of an evaluation tool also created conditions for more equitable review by applying structure and common criteria to a process previously driven by *ad hoc* judgment and select metrics. As Row 7 in Table 3 describes, rubrics encoded a value system about what faculty routinely wanted to see in students and reduced the power of the GRE as a single factor. Creating rubrics did not come with the policy-making hurdles of changing policy, and some programs developed their first one in a single meeting. However, compromising on deliberation to expedite rubric development may risk symbolic adoption (Wang & Hsieh, 2006), which may delay institutionalization.

Awareness of Disparate Impact from the GRE

The most common structural change in participating PhD programs was to eliminate GRE requirements. Such change followed deliberations about the legitimacy of scores as data points and the racially and socioeconomically inequitable consequences of relying on the exam. This finding maps on to Rows 1 and 2 in Table 3, which relate to delegitimizing preferences that privilege advantaged groups (Lamont, 2012) and becoming aware of the disparate impact of typical preferences as mechanisms for interrupting inequities (Devin et al., 2017; Lewandowsky et al., 2012). In addition to removing a major barrier to minoritized groups' admissibility, participants noted that changing GRE policy marked an important symbolic shift, given the racialized history of standardized tests (Lemann, 2000). Participants also spoke about the benefits of eliminating a financial barrier to application and communicating the social commitments of their program.

We learned from these PhD programs that changing GRE policy presents a relatively high bar for change: it required dialogue, the conditions necessary for a vote, and often, coordination with other organizational entities and their policies. Our project's external evaluation nevertheless revealed a significant increase in the percent of programs that made submitting GRE scores optional or stopped collecting them: 38% of programs reported not using/collecting the GRE at baseline vs. 82% after participation ($p=0.073$). Leaders solidified new policies through strategies that included internal assessment to establish local legitimacy.

Implications for Practice

As more educational institutions transition to holistic review (Barcelo, 2021; Langin, 2019) and seek strategies for developing and sustaining diversity without race-conscious admissions as a tool, we highlight four implications for practice that follow from our findings.

Empower change agents with racial literacy in the transition to holistic review

Participants in our study were arguably successful in implementing holistic review. Relative to other mechanisms for disrupting institutionalized inequalities, however, we found less evidence that they actively surfaced or challenged race and racialized assumptions (Table 3, Row 3). California's affirmative action ban and the purposefully broad framing of questions in our interview protocol questions may help explain this pattern in our data. The widespread use of loosely defined terms such as diversity, inclusion, and even equity, may also undermine clarity about addressing racial inequities.

Regardless of the reasons, our data suggest that when there are prohibitions on directly considering race as a factor in admissions, that race neutrality may edge into race evasiveness. Race evasiveness is critical to track because it can perpetuate racial bias and racial inequities in selection practices and policies that are race-neutral on the surface (Liera and Hernandez, 2021).

To avoid this risk, professional development and other learning opportunities should equip leaders not only to change practices that produce racially disparate outcomes, but also to challenge racially biased language, assumptions, and judgments that show up in policy and process. Even in ostensibly race-neutral processes, the interpretive machinations through which people make racialized judgments may institutionalize inequality (Lamont, 2009; Bensimon, et al., 2016; Posselt, 2020). To disrupt them, professional development should empower faculty to discuss racism, racialized organizational cultures, and implications for policy and practice.

Reimagine merit

Our findings imply a related need to continue the challenging cultural work of reframing what constitutes merit and who belongs in doctoral education. We saw that one step toward developing alternative perspectives is developing a routine of contextualizing information that has been taken at face value. Indeed, a routine of contextualization can improve any evaluation in education settings, including grading, writing letters of recommendation, peer review, and hiring. How people interpret information to develop judgments of excellence is perhaps the most fundamental part of evaluation (Lamont, 2012), because meanings associated with common criteria are deeply ingrained, sometimes at the level of one's own identity (Lamont, 2009; Posselt, 2016). Without new scripts for interpreting information, people may continue delivering racialized judgments via holistic review. It takes time, reinforcement, and practice to unlearn stereotypes and mental schema, however (Devine et al, 2017; Forscher et al., 2019; Grbic et al., 2019). In graduate admissions, off-cycle opportunities for programs to engage in what the ASA called "self-scrutiny" (ASA, 2021) may be critical to meaningful change, particularly if graduate students are themselves included in collective reflection and dialogue.

Provide support for redesigning admissions

A related implication is that faculty need time and support to change processes that involve cultural assumptions and political dynamics. For changes to stick, they cannot be expected to occur overnight. People leading the implementation of holistic admissions in our study benefited from the support of department leaders and incentives from graduate school leadership. Positive examples of such support exist: The University of Michigan Rackham Graduate School offers modest summer funding for faculty to revamp their admissions policies without the GRE. The University of California, Berkeley Graduate Division sponsored a call for PhD programs to develop proposals to improve program climate and mentorship and to redesign admissions and recruitment. They also offered professional development for program leaders. Such incentives ensure that the labor required to carry out change does not add to uncompensated racialized and gendered service burdens.

Revisit routines at multiple levels

As organizations undertake change work, our findings highlight the value of admissions committees systematically revisiting routines at multiple levels: formal policy, shared practice, and individual interpretation. Layered and coordinated change work that touches the structural, cultural, and cognitive dimensions of admissions is more likely to disrupt inequities than altering individual bias alone. We cannot count on professors bringing their best thinking to admissions, which typically occurs in conditions that elicit implicit bias (Klitgaard, 1985; Kahneman, 2011; Milkman et al., 2015). Policy and practice routines provide a backstop to biases in individuals' thinking and behavior.

Implications for Policy

Our findings also have implications for public and institutional policies concerning race and admissions. As may be the case with admissions decision makers navigating the legal landscape created by the SFFA decisions, faculty in our study sought concrete data and guidance about what is and is not permissible. Higher education institutional policy and public policy alike should be clear not only about what actions are prohibited, but also what individual and organizational actions are lawful and advised. Chief Justice John Roberts provided an example of such guidance in asserting, “Nothing in this opinion should be construed as prohibiting universities from considering an applicant’s discussion of how race affected his or her life, be it through discrimination, inspiration, or otherwise” (*SFFA vs. President and Fellows of Harvard College*, 2023).

Clear guidance is necessary because public policy creates conditions in which routines of practice are performed, and legal prohibitions on affirmative action create ambiguity about how decision-makers may discuss race (Poon et al., 2020). Combined with fear of legal liability, this ambiguity can repress any discussion of race or racism (Garces et al., 2021; Garces & Bilyalov, 2019). It may prevent people using holistic admissions from 1) speaking up against racially biased assumptions, comments, and judgments when they emerge and 2) accounting for race in ways that are lawful. The divergence of race-neutral policy from everyday racialized practices (which are legitimized by claims of policies’ race neutrality) is a mechanism through which organizations produce unequal outcomes (Ray, 2019). Policy that clarifies what *is permitted* may empower professionals to navigate legal complexities with confidence.

Implications for Research

Future research should examine evaluation, decision making, and outcomes in other educational settings that have implemented models of holistic review. How do faculty interpret test scores differently in test-optional vs. test-requiring graduate programs? When social-emotional competencies are formally assessed, which ones are most frequently a focus? Do professors interpret them from asset and/or deficit-based perspectives? How do faculty read diversity contributions from different applicants? Racialization in STEM culture is a root cause of inequality, and research should continue to investigate the default values in whatever model of “meritocracy” predominates. Across the settings where holistic review is practiced, such questions have empirical value and speak to larger issues of culture change that have been acknowledged as critical to the future of higher education (e.g., NASEM, 2018).

Rubrics’ design, use, and impacts under different conditions also need research. Researchers could measure, for example, how rubric-based review affects outcomes in organizations with different degrees of structure and centralization, using cohort-based vs. direct admit models, with more and fewer applicants, and with different criteria or weighting schemes. Researchers may also consider how faculty in different disciplines respond to, use, and refine rubrics, in admissions and for other types of evaluations. What goes into evaluation protocols, how users interpret criteria, and how rubrics are incorporated into decision making processes affect their potential (Blair-Loy et al., 2022). Rubrics are a technology, so their design and implementation likely matter for effecting change.

Finally, our findings support the view that if we do not continually assess the alignment of policies, values, practices, and outcomes, then new routines can become the latest technology in the production of inequality (Ray, 2019). Whether sensitivity to context induced by COVID-

19 will stick and whether how the Supreme Court's recent judgments affect policy, practice, and diversity outcomes all remain to be seen. Regardless of the legal context for race-conscious admissions, it is critical for organizations working to improve admissions and hiring policies to disaggregate outcomes at multiple stages (e.g., application, interviews, offers) by race/ethnicity, gender, and other characteristics of interest. Given that faculty control admissions, their awareness of such data is critical. In STEM fields, where faculty prize data at a deep cultural level, ongoing assessment may have special salience. We saw that the legitimacy of new routines may hinge upon local data under new policy and leaders' ability to translate that data. For scientists, after all, holistic review is effectively an experiment. And as with any move from experiment to theory development, changing scientists' views about the best approach to admissions may involve not only data analysis, but also a lively debate about its interpretation.

References

- American Sociological Association. (2021). Reconsidering the Use of the Graduate Record Examination (GRE) in Graduate School Admissions Decisions.
https://www.asanet.org/sites/default/files/gre_statement_march_2021-final_format.pdf
- Association of American Medical Colleges. (2010). Roadmap to diversity: Integrating holistic review practices into medical school admission processes. AAMC.
- Association of American Medical Colleges. (2014). Roadmap to diversity and educational excellence. Association of American Medical Colleges.
- Attiyeh, G., & Attiyeh, R. (1997). Testing for bias in graduate school admissions. *Journal of Human Resources*, 524-548.
- Baker, R., Klasik, D., & Reardon, S. F. (2018). Race and stratification in college enrollment over time. *AERA Open*, 4(1), 2332858417751896.
- Bal, A., Afacan, K., & Cakir, H. I. (2018). Culturally responsive school discipline: Implementing learning lab at a high school for systemic transformation. *American Educational Research Journal*, 55(5), 1007-1050.
- Ball, D. L. (2018, July 16). (What) does practice-based teacher education have to do with disrupting racism and oppression? AERA Presidential Talk.
- Barceló, N. E., Shadravan, S., Wells, C. R., Goodsmith, N., Tarrant, B., Shaddox, T., ... & DeBonis, K. (2021). Reimagining Merit and Representation. *Academic Psychiatry*, 1-9.
- Barton, K. (2015). Elicitation techniques: Getting people to talk about ideas they don't usually talk about. *Theory and Research in Social Education*, 43(2), 179–205.
- Bastedo, M., & Jaquette, O. (2011). Running in place: Low-income students and the dynamics of higher education stratification. *Educational Evaluation and Policy Analysis*, 33(3).

- Bastedo, M. N., Bowman, N. A., Glasener, K. M., & Kelly, J. L. (2018). What are we talking about when we talk about holistic review? Selective college admissions and its effects on low-SES students. *Journal of Higher Education*, 89(5), 782-805.
- Belasco, A. S., Rosinger, K. O., & Hearn, J. C. (2015). The test-optional movement at America's selective liberal arts colleges: A boon for equity or something else? *Educational Evaluation and Policy Analysis*, 37(2), 206-223.
- Bennett, C. T. (2021). Untested admissions: Examining changes in application behaviors and student demographics under test-optional policies. *American Educational Research Journal*, 00028312211003526.
- Bensimon, E. M., Dowd, A. C., & Witham, K. (2016). *Five principles for enacting equity by design*. Diversity and Democracy, The Equity Imperative. AACU. (19)1.
- Bielby, R., Posselt, J. R., Jaquette, O., & Bastedo, M. N. (2014). Why are women underrepresented in elite colleges and universities? A non-linear decomposition analysis. *Research in Higher Education*, 55, 735-760.
- Blair-Loy, M., Mayorova, O. V., Cosman, P. C., & Fraley, S. I. (2022). Can rubrics combat gender bias in faculty hiring?. *Science*, 377(6601), 35-37.
- Bourdieu, P. (1987). *Distinction: A social critique of the judgment of taste*. Harvard Univ Press.
- Bourdieu, P., & Passeron, J. (1990). *Reproduction in education, society and culture* (V. 4). Sage.
- Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, 5(2), eaau4734.
- Charmaz, K. (2014). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage Publications.

- Kent, J. D., & McCarthy, M. T. (2016). Holistic review in graduate admissions. Council of Graduate Schools.
- Cheryan, S., & Markus, H. R. (2020). Masculine defaults: Identifying and mitigating hidden cultural biases. *Psychological Review*, 127(6), 1022.
- Creswell, J. (2009). *Research design: Qualitative, quantitative, mixed methods approaches*. Sage.
- D'adderio, L. (2008). The performativity of routines: Theorising the influence of artefacts and distributed agencies on routines dynamics. *Research Policy*, 37(5), 769-789.
- Devine, P. G., Forscher, P. S., Cox, W. T. L., Anna Kaatz, Sheridan, J., Carnes, M. (2017). A gender habit-breaking intervention led to increased hiring of female faculty in STEM departments. *Journal of Experimental Social Psychology*, 73, 211-215.
- Diamond, J., & Lewis, A. (2019). Race and discipline at a racially mixed high school: Status, capital, and the practice of organizational routines. *Urban Education*, 54(6), 831-859.
- Eaton, A., Saunders, J., Jacobson, R., & West, K. (2019). How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates. *Sex Roles*, 82(3-4), 127-141.
- Educational Testing Service. (2015). Avoid these five common mistakes when using GRE scores https://www.ets.org/s/gre/pdf/infographic_5_mistakes.pdf
- Feldman, M. S. (2000). Organizational routines as a source of continuous change. *Organization Science*, 11(6), 611-629.
- Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48(1), 94-118.

- Feldman, J. (2018). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin Press.
- Feingold, A. (1994). Gender differences in personality. *Psychological Bulletin*, 116(3), 429.
- Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five US racial groups. *Personnel Psychology*, 61(3), 579-616.
- Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T. L., Devine, P. G. (2017). Breaking the prejudice habit. *Journal of Experimental Social Psychology*, 72, 133-146.
- Garces, L. (2012). Racial diversity, legitimacy, and the citizenry: The impact of affirmative action bans on graduate school enrollment. *Review of Higher Education*, 36(1), 93-132.
- Garces, L. M. (2013). Understanding the impact of affirmative action bans in different graduate fields of study. *American Educational Research Journal*, 50(2), 251-284.
- Garces, L. M. (2014). Aligning diversity, quality, and equity: The implications of legal and public policy developments for promoting racial diversity in graduate studies. *American Journal of Education*, 120(4), 457-480.
- Garces, L. M., & Bilyalov, D. (2019). Navigating the quicksand: How postsecondary administrators understand the influence of affirmative action developments on racial diversity work. *Teachers College Record*, 121(3), 1-40.
- Garces, L. M., Johnson, B. D., Ambriz, E., & Bradley, D. (2021). Repressive Legalism: How Postsecondary Administrators' Responses to On-Campus Hate Speech Undermine a Focus on Inclusion. *American Educational Research Journal*, 58(5), 1032-1069.
- Grbic, Morrison, E., Sondheimer, H. M., Conrad, S. S., & Milem, J. F. (2019). The Association Between a Holistic Review in Admissions and the Diversity of Accepted Applicants and Students Matriculating to Medical School. *Academic Medicine*, 94(3), 396-403.

- Hall, S., & Morley, D. (2019). *Essential Essays, Volume 1: Foundations of Cultural Studies*. Durham: Duke University Press.
- Harper, S. R. (2010). An anti-deficit achievement framework for research on students of color in STEM. *New Directions for Institutional Research*, 2010(148), 63-74.
- Hirschman, D., Berrey, E., & Rose-Greenland, F. (2016). Dequantifying diversity: Affirmative action and admissions at the University of Michigan. *Theory and Society*, 45(3), 265-301.
- Hirschman, D., & Bosk, E. A. (2020). Standardizing biases: Selection devices and the quantification of race. *Sociology of Race and Ethnicity*, 6(3), 348–364.
- Holland, M. M., & Ford, K. S. (2020). Legitimizing prestige through diversity. *The Journal of Higher Education*, 1-30.
- Hsee, C. (1995). Elastic justification in decision making: How task-irrelevant but tempting factors influence decisions. *Organizational Behavior and Human Decision*, 62, 330–337.
- Johnson, J. C., & Weller, S. C. (2002). Elicitation techniques for interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 491–514). Sage Publications.
- Jones, A., Schuhmann, P., Soques, D., & Witman, A. (2020). So you want to go to graduate school? Factors that influence admissions to economics PhD programs. *Journal of Economic Education* 51(2), 177-190.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Karen, D. (1990). Toward a political-organizational model of gatekeeping: The case of elite colleges. *Sociology of Education*, 227-240.
- Klitgaard, R. (1985). *Choosing Elites*. Basic Books.

- Kyllonen, P. C. (2005). The Case for Noncognitive Assessments. ETS R&D Connections. Downloaded from https://www.ets.org/Media/Research/pdf/RD_Connections3.pdf;
- Kyllonen, P., Walters, A. M., & Kaufman, J. C. (2005). Noncognitive constructs and their assessment in graduate education: A review. *Educational Assessment*, 10(3), 153-184.
- Kyllonen, P., Walters, A., & Kaufman, J. (2011). The role of noncognitive constructs and other background variables in graduate education. ETS Research Report Series, 2011(1), i-133.
- Lamont, M. (2009). *How professors think*. Harvard University Press.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38, 201-221.
- Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. Macmillan.
- Liera, R., & Hernandez, T. E. (2021). Color-evasive racism in the final stage of faculty searches: Examining search committee hiring practices that jeopardize racial equity policy. *The Review of Higher Education*, 45(2), 181-209.
- Lipsky, M. (1980). *Street level bureaucracy: Dilemmas of the individual in public services*. Sage.
- Langin, K. (2019). Ph.D. programs drop standardized exam: Amid concerns about diversity and the test's utility, many are moving away from the GRE. *Science* 364 (6443).
- Lewandowsky, Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction. *Psychological Science in the Public Interest*, 13(3), 106–131.
- Maitlis, S., & Sonenshein, S. (2010). Sensemaking in crisis and change: Inspiration and insights from Weick (1988). *Journal of Management Studies*, 47(3), 551-580.
- Maxwell, J. A. (2013). *Qualitative research design*. Sage Publications.
- McGee, E. O. (2020). Interrogating structural racism in STEM higher education. *Educational Researcher* 49 (9), 633-644.

- McNair, T., Bensimon, E. M., & Malcom-Piqueux, L. (2020). *From equity talk to equity walk: Expanding practitioner knowledge for racial justice in higher education*. Wiley.
- Milkman, K. L., Rogers, T., & Bazerman, M. H. (2008). Harnessing Our Inner Angels and Demons. *Perspectives on Psychological Science*, 3(4), 324–338
- Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, 100(6), 1678.
- Miller, C., & Posselt, J. (2020). Equitable admissions in the time of COVID-19. *Physics 13*, 199.
- Miller, C., & Stassun, K. (2014). A test that fails. *Nature*, 510 (7504), 303-304.
- Miller, C. W., Zwickl, B. M., Posselt, J. R., Silvestrini, R. T., & Hodapp, T. (2019). Typical physics Ph. D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion. *Science Advances*, 5(1), eaat7550.
- Miller, C. W., Zwickl, B. M., Posselt, J. R., Silvestrini, R. T., & Hodapp, T. (2020). Response to comment on “Typical physics Ph. D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion”. *Science Advances*, 6(23), eaba4647.
- National Academies of Science, Engineering, and Medicine. (2018). *Graduate STEM Education for the 21st Century*. The National Academies Press. <https://doi.org/10.17226/25038>
- Ortiz, A. M. (2003). The ethnographic interview. In F. K. Stage & K. Manning (Eds.), *Research in the college context: Approaches and Methods* (pp. 35-48). Brunner-Routledge.
- Oswald, F. L., & Hough, L. M. (2011). Personality and its assessment in organizations: Theoretical and empirical developments. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, V2*. (pp. 153–184). American Psychological Association.

- Pentland, B., & Feldman, M. (2005). Organizational routines as a unit of analysis. *Industrial and Corporate Change*, 14(5), 793–815.
- Petersen, S. L., Erenrich, E. S., Levine, D. L., Vigoreaux, J., & Gile, K. (2018). Multi-institutional study of GRE scores as predictors of STEM PhD degree completion. *PLoS one*, 13(10), e0206570.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61(1), 153-172.
- Poon, O., Lee, D., Galvez, E., Serrano, B. C., Raza, A., Chun, N., Engler, J. S. (2020). Race-conscious and unconscious holistic admissions. Paper presentation at Association for the Student of Higher Education Annual Conference, New Orleans, LA.
- Posselt, J. R. (2015). Disciplinary logics in doctoral admissions: Understanding patterns of faculty evaluation. *The Journal of Higher Education*, 86(6), 807-833.
- Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press.
- Posselt, J. R. (2020). *Equity in science: Representation, culture, and the dynamics of change in graduate education*. Stanford University Press.
- Posselt, J. R., Jaquette, O., Bielby, R., & Bastedo, M. N. (2012). Access without equity: Longitudinal analyses of institutional stratification by race and ethnicity, 1972–2004. *American Educational Research Journal*, 49(6), 1074-1111.
- Posselt, J. R., Jaquette, O., Bielby, R., & Bastedo, M. N. (2012). Access without equity: Longitudinal analyses of institutional stratification by race and ethnicity, 1972–2004. *American Educational Research Journal*, 49(6), 1074-1111.

- Posselt, J., Hernandez, T. E., Villarreal, C. D., Rodgers, A. J., & Irwin, L. N. (2020). Evaluation and decision making in higher education: Toward equitable repertoires of faculty practice. *Higher Education: Handbook of Theory and Research: Volume 35*, 1-63.
- Powell, W., & DiMaggio, P. (1991). *The new institutionalism in organizational analysis*. University of Chicago Press.
- Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3), 375–392.
- Ray, R., & Brown, J. (2015). Reassessing student potential for medical school success: distance traveled, grit, and hardiness. *Military Medicine*, 180(suppl_4), 138-141.
- Ray, V. (2019). A theory of racialized organizations. *American Sociological Review* 84(1), 26-53.
- Ray, V., & Purifoy, D. (2019). The colorblind organization: Race, organizations, and the organizing process. *Research in the Sociology of Organizations*, 60, 131-150.
- Rodgers, A. (2020). Toward a racialized cultural historical activity theory in higher education. Paper presentation at American Educational Research Association, San Francisco, CA.
- Rosinger, K., Sarita Ford, K., & Choi, J. (2021). The role of selective college admissions criteria in interrupting or reproducing racial and economic inequities. *Journal of Higher Education*, 92(1), 31-55.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist*, 56(4), 302.
- Saldaña, J. (2013). *The coding manual for qualitative researchers*. Sage Publications.
- Sedlacek, W., & Brooks, G.(1976). *Racism in American education*. Chicago: Nelson-Hall.

- Sewell Jr, W. H. (1992). A theory of structure: Duality, agency, and transformation. *American Journal of Sociology*, 98(1), 1-29.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin*, 124(2), 262.
- Slay, K. (2017). Choosing colleges in a post-affirmative action era: Black students' perceptions of institutional diversity and campus climate (Doctoral dissertation, Univ of Michigan).
- Spillane, J. P., Hopkins, M., & Sweet, T. M. (2018). School district educational infrastructure and change at scale. *American Educational Research Journal*, 55(3), 532-571.
- Stevens (2007). *Creating a class*. Harvard University Press.
- Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). Sage.
- Students for Fair Admissions vs. President and Fellows of Harvard College*. (2023). 600 U. S. 39.
- Thornhill, T. (2019). We want Black students, just not you: How White admissions counselors screen Black prospective students. *Sociology of Race and Ethnicity*, 5(4), 456–470.
- Wang, W., & Hsieh, J. J. (2006). Beyond routine: Symbolic adoption, extended use, and emergent use of complex information systems in the mandatory organizational context. ICIS 2006 Proceedings. Paper 48. <http://aisel.aisnet.org/icis2006/48>
- Verostek, M., Miller, C. W., & Zwickl, B. (2021). Analyzing admissions metrics as predictors of graduate GPA and whether graduate GPA mediates Ph.D. completion. *Physical Review Physics Education Research*, 17(2), 020115.
- Warikoo, N. (2018). What meritocracy means to its winners: Admissions, race, and inequality at elite universities in the United States and Britain. *Social Sciences*, 7(8), 131.

Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies*. Simon and Schuster.

Wilson, M., DePass, A., & Bean, A. (2018). Institutional interventions remove barriers to recruit and retain diverse biomedical PhD students. *CBE Life Sciences Education* 17(2), ar27.

W.T. Grant Foundation. (2021). Research Practice Partnerships.

<https://rpp.wtgrantfoundation.org/>

Young, N. T., & Caballero, M. D. (2021). Physics Graduate Record Exam does not help applicants “stand out.” *Physical Review Physics Education Research*, 17(1), w010144.