
Toward Inclusive Excellence in Graduate Education: Constructing Merit and Diversity in PhD Admissions

JULIE R. POSSELT
University of Michigan

Professors play an underexamined role as gatekeepers, and their understandings of merit have significant implications for racial equity and diversity in graduate education and the professoriate. To understand faculty reliance upon admissions criteria that undermine espoused diversity goals, this study examined decision making in 10 highly selective doctoral programs, including the meanings faculty associate with common admissions criteria. Through 86 interviews and 22 hours of admissions committee observations, findings reveal that conceptions of merit changed throughout the review process. Privileging diversity among those who made the short list marginally affected outcomes because the initial standard—a very high quantitative bar of conventional achievement—excluded many students of color. Implications for reframing merit and reforming graduate admissions are discussed.

Vivek (committee chair): He grew up in a yurt in the Himalayas, was raised by his mom and grandma after his father died at an early age, and the next neighbors were two mountains over. He then found his way to a major US research university and has since started the only organization for the discipline in the Himalayan region.

William (assistant professor): But do we think he can succeed? [*long pause*]

Vivek: He's the most amazing case we've ever seen.

Harold (professor): He would bring some personality to the department. I commit to look after him and fund him through the prelims. . . . He presents himself as quite intelligent.

Ryan (graduate student): Excellent idea to give him a chance.¹

Electronically published June 3, 2014

American Journal of Education 120 (August 2014)

© 2014 by The University of Chicago. All rights reserved.

0195-6744/2014/12004-0002\$10.00

Introduction

Professors play an underexamined role as gatekeepers of the professions, including the professoriate. One context in which this gatekeeping occurs is admission into graduate programs, which entails evaluative processes that are often opaque to outsiders and taken for granted by insiders (Klitgaard 1985). The lack of transparency about what holistic review entails may compromise public perceptions of its fairness, driving some to assume that race plays a pivotal role in outcomes (Guinier 2003) instead of serving as a “factor of a factor of a factor,” as Justice Ruth Bader Ginsburg noted in her 2013 dissenting opinion on *Fisher v. University of Texas* (2013, 3).

This article pulls back the curtain on holistic review through ethnographic research. In particular, I focused on the ways that faculty operationalize, assess, and relate two interests: *merit*, the principle of desert for allocating opportunities, and *diversity*, which may or may not be conceived of as part of merit and may or may not include considerations about race and ethnicity (Klitgaard 1985). Joining merit and diversity in an integrated notion of inclusive excellence is “critical to the wellbeing of democratic culture” (Association of American Colleges and Universities 2013, para. 2) and enables postsecondary institutions to “leverage diversity for student learning and institutional excellence” (Milem et al. 2005, v). By examining the relationship of merit and diversity in elite graduate programs, I therefore sought to clarify the extent of progress needed toward a culture of inclusive excellence.

Results of 2 years of fieldwork, including 86 interviews and 22 hours of admissions committee meeting observations, reveal that diversity is one interest that faculty pursue, but that it is secondary to conventional, quantitative notions of merit—both in overall importance and temporally in the evaluative process. Faculty members in the highly selective programs in this study feel that evaluating doctoral student applications is professionally, cognitively, and politically difficult work, for it involves balancing a range of criteria and organizational interests. What is more, evaluations happen in a context of incomplete information, limited time, and personal and collective standards of judgment. In this balancing act, faculty conceptualize diversity in mostly pragmatic terms; moreover, what counts as merit only includes diversity when judging among a handful of borderline applicants.

JULIE R. POSSELT is an assistant professor of higher education at the University of Michigan. Her research examines institutionalized inequalities in higher education and efforts to resist these historic tendencies and encourage diversity. She has particular interests in graduate education, admissions, research methods, and organizational theory.

Literature Review: Defining Merit and Diversity

The ideal of merit drives academic evaluations and academia, more broadly (Lamont 2009, 242), but what constitutes merit is socially constructed, contextualized, and contested (Espeland and Sauder 2007; Stevens 2008). Most research on graduate admissions implicitly paints merit as an individual good—a characteristic that inheres in applicants to varying degrees according to possession of specific characteristics. Graduate Record Exam (GRE) scores are among the very strongest predictors of admission (Sternberg and Williams 1997; Willingham 1974),² along with perceptions of professional competency, inter- and intrapersonal problems, the reputation of one's undergraduate institution, GPA, and stated religion and race/ethnicity (Attieh and Attieh 1997; Brear et al. 2008; Campbell 2009; Eide et al. 1998; Gartner 1986).

However, research on undergraduate admissions teaches us that merit is not only an individual characteristic but also an organizational challenge (Kahlenberg 1996; Karabel 2005; Karen 1990). As in hiring, admissions decision makers approach selection to realize their missions and maximize organizational interests (Birnbaum 1988; Guinier 2002; Klitgaard 1985), not only to identify applicants who are likely to succeed. Undergraduate and medical admissions officers, for example, balance individual applicant qualifications with collective concerns about prestige, diversity, fiscal viability, and the desired profile of a cohort (Killgore 2009; Nivet 2011; Stevens 2008).

As such, reviewers do not assess individuals in an absolute sense, but rather comparatively evaluate them against others in the pool, against current students, and even against themselves (Stevens 2008). These social and organizational dynamics of merit have been a focus of undergraduate admissions research since the 1970s, when Wechsler (1977) argued, “the essence of selective admissions was the subjective judgment of the admissions officer” (244). To my knowledge, the current study is the first to bring an organizational perspective on merit to the study of graduate admissions. I consider merit as mutually constituted by the qualities of students, the judgment of faculty who evaluate them, and constraints that the review process imposes.

Operationally, diversity simply means heterogeneity on a variety of characteristics, but beyond the “laundry list” approach, it “represents a powerful change in the environment . . . an imperative that must be engaged if institutions are to be successful in a society that is ever more pluralistic” (Smith 2009, 49). Given America's increasing pluralism, scholars, administrators, and courts alike uphold racial diversity as integral to merit (Guinier 2003; Gurin et al. 2004; Milem et al. 2005; Orfield 2001) because of the educational benefits it confers (e.g., Antonio 2004; Denson and Chang 2009; *Grutter et al. v. Bollinger et al.* 2003; Gurin et al. 2002; Hurtado 2001).³ Doctoral education possesses

Toward Inclusive Excellence in Graduate Education

a special warrant for gatekeeping on the basis of diversity, as the primary training ground for knowledge production in an intellectual context of increasing epistemological diversity (Page 2008; Pallas 2001). Ralph Waldo Emerson (1875) wrote, “Our knowledge is the amassed thought and experience of innumerable minds” (200), and contemporary knowledge production demands cognitive complexity that diverse teams of scholars proffer (Neumann 1991). Diversity in higher education is also tied into democratic imperatives of equal opportunity (Chang 2002; National Center for Institutional Diversity 2011). Participants in this study raised particular concerns about racial diversity and gender equity, which reflect a broader equity challenge in graduate education, as exemplified in recent doctoral degree attainment, relative to the US population (see fig. 1).

What explains unequal doctoral enrollment and attainment? Recent research explores GRE score reliance and public policy restricting affirmative action (Garces 2011; Roach 2005). Other mechanisms, such as faculty in-group norms in gatekeeping (Campbell 2009), may also shed light on these persistent inequalities. Given underrepresented groups’ lower average GRE scores (e.g., Miller 2013), we need to know how faculty make sense of these scores and justify their continued reliance upon them in spite of the negative consequences for diversity. Therefore, I examine in this article the social construction of merit and its implications for diversity by analyzing the shared meanings that faculty attribute to graduate admissions criteria.

Most research on graduate admissions is quantitative and analyzes criteria in relation to final admissions outcomes. By contrast, I used ethnographic fieldwork to consider how faculty prioritize different considerations at different points in the review process and the tacit organizational norms and values revealed in their deliberations. The research questions guiding analysis were as follows:

1. What are the evaluative scripts of merit that faculty use in order to distinguish among applicants? How is merit operationalized?
2. How do faculty conceptualize diversity? What meanings do they attribute to diversity?
3. How do faculty relate merit and diversity? What is the place of diversity in merit conceptualizations?

Theoretical Framework

Graduate programs’ autonomy produces a wide variety of approaches to admissions, with different programs (and the same programs over time) employing models of merit that they deem suitable for the current context (Klitgaard 1985).

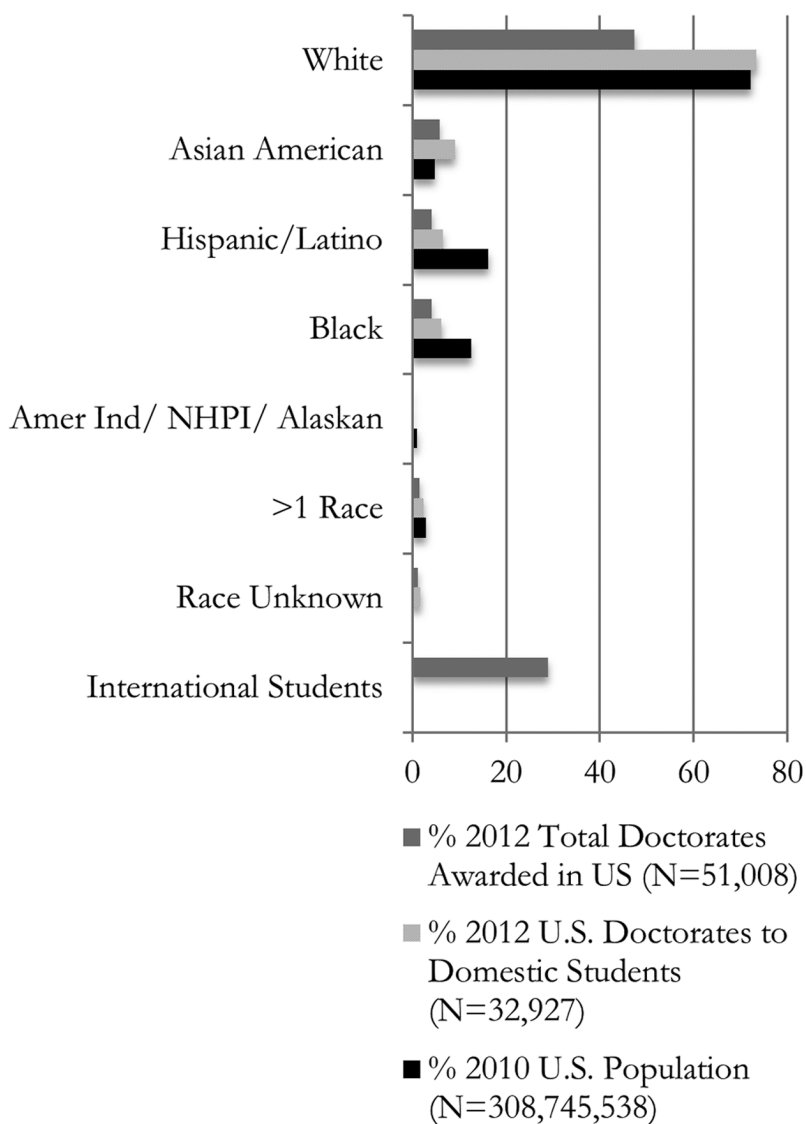


FIG. 1.—Distributions of 2012 doctorates awarded and US population, by race/ethnicity.

Toward Inclusive Excellence in Graduate Education

Below I explain how sociological and organizational research can inform an understanding of merit's construction as a matter of subjective faculty judgment. This framework includes three complementary perspectives: (a) admissions preferences as a matter of cultural taste, (b) the use of evaluative scripts to translate information about applicants into value judgments, and (c) organizational dynamics through which faculty exercise judgment.⁴

Cultural Preferences in Academic Selection

Pierre Bourdieu's work includes a theory of selection and social reproduction rooted in cultural tastes, or "manifested preferences" (Bourdieu 1984, 56). His theory does not delve into the possibilities for social change that come from nondominant groups' cultural assets and resistance of the status quo (Carter 2008; Espino 2014; Yosso 2005) but rather the social processes by which elites try to control the terms of access to opportunity in order to maintain their privileged status. I explain the contours of his framework before summarizing applications and critiques.

First, recognizing the power of academic selection in credential-based systems of social positioning (i.e., fields of interaction), Bourdieu (1977) assumes economic elites allocate educational opportunity to maintain their own privilege.⁵ It would, of course, be illegitimate to explicitly deny opportunities to low-status individuals; therefore, elites do so indirectly through symbolic exclusion. Progress up the academic hierarchy employs subjective evaluations that privilege a narrow set of social ties and cultural knowledge and behavior (i.e., social and cultural capital) that are disproportionately found among dominant groups in society (Bourdieu 1986).⁶

Over time, taste for these markers of apparent distinction becomes embedded in the common sense of a field's elite members. No longer recognized as self-serving preferences, they become legitimized as part of the field's self-evident rules and norms (i.e., *doxa*; Bourdieu 1984). For example, in *Homo Academicus*, Bourdieu (1988) reasons that those who display the greatest ease with academic work and norms (i.e., what he calls academic cultural capital) are regarded as the most brilliant. And because brilliance is the sine qua non of praise in the academy, their academic cultural capital is perceived as a legitimate basis on which to allocate opportunity.

Thinking of American society as a field of interaction in which adolescents and their families seek status and of selective colleges and universities as a field characterized by competitions for prestige, scholars have applied Bourdieu's theory to research on undergraduate admissions. They clarify how institutional power plays lead elite universities to value applicant traits that are disproportionately found among the already privileged (e.g., Karabel 2005;

Karen 1990; Stampnitzky 2006; Stevens 2008). However, in the United States, the boundaries between subcultures are often fuzzy or overlapping (Lamont 1992), and, as Stevens (2008) discusses, opportunities and institutionally valued cultural objects are defined by intersections of race, class, and gender rather than by class alone (Yosso 2005). Also, education in the United States is correlated with lower levels of cultural exclusivity and higher political tolerance, suggesting American economic elites may have broader cultural tastes than those Bourdieu identified in France (Bourdieu 1984; Bryson 1996). Recent cultural sociologies suggest that American omnivorism may extend to academic selection, with conceptions of academic merit that privilege diversity rather than regarding merit and diversity as a trade-off (Lamont 2009; Stevens 2008). The current research explores how diversity is weighed in evaluations for admission to graduate education.

Identity, Boundaries, and Scripts in Evaluation

In response to these differences by national context, Lamont's (1992) comparative study of class structure in France and the United States broadens reproduction-inducing mechanisms from status struggle to include the drawing and strength of identity-based boundaries. Tastes define group boundaries, she argues, and selection can reify them: "Exclusion is often the unintended consequence or latent effect of the definition by the upper middle class of its values and indirectly of its group identity and its nature as a community. . . . Only when boundaries are widely agreed upon (i.e., only when people agree that some traits are better than others) can symbolic boundaries take on a widely constraining (or structural) character and pattern social interaction in an important way" (Lamont 1992, 178).

This group definition function of gatekeeping may explain the importance of "fit" and "match" in admissions and hiring, often over and above qualifications (Birnbaum 1988; Brink 1999; Klitgaard 1985; Twombly 1992). Specifically, graduate education's mission of knowledge advancement (Gumport 1993a, 1993b) may elevate assessments of intellectual fit involving research interests and perceived ability. It may also elevate the importance of selecting students with diverse knowledge standpoints, given the contributions that graduate students make to the intellectual lives of academic departments (Walker et al. 2007).

Graduate education's goals are thus intertwined with how faculty make meaning of information in applications and transform those meanings into preferences. Lamont's (2009) recent work clarifies these relationships, applying the ideas about boundaries to faculty evaluation on interdisciplinary fellowship review panels. She proposes that faculty operationalize excellence and ratio-

Toward Inclusive Excellence in Graduate Education

nalize their judgments using evaluative scripts. The idea traces to Goffman's (1959) dramaturgical theory, which uses such elements as scripts, roles, performance, and the stage to illuminate social behavior. Briefly, public (or front stage) behavior can be likened to performance of familiar roles and scripts, and it occurs in response to actors' definition of a given situation and its normative demands (Goffman 1959). Following conventional social scripts also ensures that others will find the behavior (i.e., the performance) realistic and acceptable. From this angle, norm conformity involves as much impression management and performance as personal commitment.

Returning to the context of faculty judgment, Lamont (2009) proposes that faculty construct excellence by linking into evaluative scripts the criteria of evaluation and the meanings those criteria hold in light of reviewers' identities and work. Evaluative scripts thus serve as decision pathways and stories that reviewers tell to justify their judgments (Lamont 2009).⁷ Following these pathways is a bit like peeling back the layers of an onion. The surface-level criterion has meaning, but there are reasons—often implicit—that those meanings are themselves important (i.e., the meanings have meaning). For example, some may interpret grades as academic ability, ability as a critical signal of future success, and likelihood of future success as worthiness of the investment that admissions represents. Through such scripts, excellence as a social construction is reconciled with the demands of objectivity that inhere in allocating elite opportunities, for “most reviewers uphold the legitimacy of the process by seamlessly folding their idiosyncratic preferences into the formal criteria of evaluation” (Lamont 2009, 130). Individual constructions of excellence come to constitute boundaries that define scholars and become the terms on which they define others.

An Organizational Perspective on Academic Gatekeeping

The importance of fit and match, the use of admissions to seek status, and the broader group definition function of gatekeeping illustrate organizational influences in holistic review. Evaluation outcomes are also contingent on committee relations, characteristics of the applicant pool, and the review process itself. Deliberations in hiring academic administrators often require sense-making to distinguish among plausible candidates (Birnbaum 1988; Twombly 1992). For example, by clarifying what they mean by a criterion like “professional experience,” a search committee may disqualify some candidates, see others' strengths, and build up the committee by identifying shared values. Rather than implicit criteria driving the process, as Bourdieu describes, these findings suggest implicit criteria become explicit through deliberation. Often decisions come down not to qualifications but to how one individual “rep-

resents the optimization of one or another institutional goal” (Birnbaum 1988, 498).

Two common goals that organizations pursue through their decisions are fulfilling present identity and positioning for future strength. These goals underlie decision-making logics of appropriateness and consequences, respectively, and like evaluative scripts, they clarify pathways to decisions (March 1994). Under a logic of appropriateness, decision makers assess the situation, assess their identity (or their organization’s identity), and consider what a person (or organization) with such an identity should do in such a situation (March 1994). A logic of consequences, on the other hand, demands that decision makers anticipate the future. They assess available alternatives, expected consequences, and the relative value of those consequences (March 1994). Together, evaluative scripts and decision-making logics clarify how faculty think about and use selection criteria in specific cultural contexts.

The cultural context in this research is highly ranked graduate programs. Status must be conferred by an outside party (Weber [1922] 1978), and to legitimize and maintain a high ranking in their respective disciplines, program leaders frame and order their program for maximum consistency with field-level norms (Karen 1990; March 1994). In admissions, this includes aligning their own selection criteria with those that carry the greatest cachet in the organizational field (Bourdieu and Nice 1984; Espeland and Sauder 2007). And in graduate education, those criteria include two of the strongest predictors of graduate schools admission generally: high GRE scores and selectivity of undergraduate institution (Attiyeh and Attiyeh 1997).

However, under these broad trends, theory and research also portray a highly negotiated, contingent process in which individual outcomes could hardly be predicted from expected or formal criteria. As graduate programs use selection processes to reinforce or forge new organizational identities and goals—such as disciplinary prestige, specific intellectual projects, or diversity—desirable applicant characteristics come into focus because they support the program’s vision. This perspective helps reconcile evaluators’ belief in a fair process with the reality that social matching and homophilic judgments occur (Azoulay et al. 2009). Further, it reveals how deliberative processes have the potential to either reify conventional scripts of merit or, through development of counterscripts, develop more a more inclusive vision of excellence.

Research Design

The research design is an ethnographic comparative case study of the graduate admissions cycle in 10 highly selective PhD programs across the humanities, social sciences, and natural sciences. This article reflects findings from extended

Toward Inclusive Excellence in Graduate Education

fieldwork I conducted over two admissions cycles (September to April in 2010–11 and 2011–12) in 10 PhD programs in three large research universities. In this section, I describe the sampling design, the data collection, and the analysis strategies, and some limitations of the study.

Sampling

I selected these universities for administering the type of doctoral programs in which I was most interested. Confidentiality agreements with participants and the Institutional Review Board constrain me from describing the university contexts in great detail, but two are public and one private, and two are on the West Coast of the United States and one is in the Midwest.

The sample has three levels: disciplines, departments/PhD programs,⁸ and the individuals within them. With a goal of studying up to 12 PhD programs, I narrowed my pool to programs that are ranked in the top 15 in their discipline. Many qualified individuals apply to these programs; therefore, the competing demands of selection come into sharper focus. Of the 10 programs represented in this article, four are in the humanities (classics, linguistics, and two in philosophy), three in the social sciences (political science, economics, and sociology), and three in the physical sciences (biology, physics, and astronomy/astrophysics). Within these programs, my sample in each department consists of the admissions committees, additional members of the faculty with many and few years of experience in graduate admissions, and one emeritus professor. Three programs included graduate students on the committee, whom I interviewed as well. The sample includes a total of 68 individuals (62 faculty and six graduate students), for an average of about seven per program. Table 1 summarizes demographic characteristics by program and disciplinary area.⁹

Data Collection

By analyzing interview data, I captured the meanings that participants associate with various criteria and how they interpret the information they have about applicants. I conducted 86 semi-structured interviews of about 45 minutes each with each admissions committee chair, other current faculty on the admissions committee, and one emeritus professor in each program. Where this did not elicit at least six participants, I also interviewed additional members of the program faculty with many and few years of admissions experience.

I conducted 2–4 interviews with each admissions chair, including an initial informational interview in the fall and a follow-up interview in the spring. I

TABLE 1

Sample Characteristics

Program	N	Female (%)	Born outside the United States (%)	Scholars of Color (%)	Domestic Scholars of Color (%)	Number of Graduate Students
Full sample	68	18	36	15	3	6
Humanities	25	33	26	7	4	2
Classics	5	40	40	0	0	0
Linguistics	7	29	29	14	0	2
Philosophy 1	6	33	33	0	0	0
Philosophy 2	7	29	14	14	14	0
Social sciences	24	15	35	16	4	3
Economics	6	0	33	0	0	0
Political science	8	25	38	13	0	3
Sociology	10	20	33	34	10	0
Natural sciences	19	6	46	21	0	1
Astrophysics	6	0	33	33	0	1
Biology	6	17	33	17	0	0
Physics	7	0	71	14	0	0

gave other participants the option of one longer interview or two shorter interviews—one just before and another after file review had occurred. Most opted for a single interview. We discussed participants’ own academic socialization, previous experiences with graduate admissions, important criteria and what they are perceived to signal, and how traits of highly valued faculty compare to those of compelling applicants. Using interview strategies employed by Lamont (1992, 2009) and Tierney and Bensimon (1996), I also inquired about “ideal types” of applicants to draw out the ways faculty construct merit in relation to specific individuals. With those whom I conducted follow-up interviews, we also discussed profiles of admitted and rejected students and how it came to be that those applicants were selected or not. To contextualize current practice, I interviewed an emeritus professor in each program for oral histories of developments in their departments and disciplines.

As part of his dramaturgical theory, Goffman (1959) proposes a dichotomy in human behavior between “front stage,” which we make visible for others, and “back stage,” which is, “out of bounds to members of the audience” (124). Admissions is one of many social activities with official and unofficial stories (i.e., about why applicants are admitted/rejected) and espoused and enacted values (i.e., that contribute to admission evaluations and outcomes). For these reasons, and because selection is ultimately a collective activity, I comple-

Toward Inclusive Excellence in Graduate Education

mented individual interviews with observations of admissions committee and other meetings to gather “back stage” knowledge of admissions practice. Committee interactions also provided a crucial window into cultural connections between the discipline and admissions criteria and processes. Six of the 10 programs consented to observational data collection,¹⁰ through which I composed systematic field notes and reflections about committee interactions, the frequency with which various criteria are mentioned, and the combinations of criteria that are salient for each applicant.

Data Analysis

A professional transcriptionist transcribed 75% of the interview audio files, and I transcribed the remaining 25% to facilitate reflection on my practice as an interviewer and determination of whether the protocol needed to be adapted. I used qualitative research software (NVivo version 9.2) for ongoing composition of memos, coding, and analysis of transcripts and field notes. Case study recommends processes for sampling but not data analysis (Stake 2005); thus, I employed constant comparison (Corbin and Strauss 2008; Glaser and Strauss 1967; Miles and Huberman 1994), which specifies open, axial, and selective coding as systematic stages of analysis.

More specifically, I conducted line-by-line analysis of each interview transcript, adding codes as new themes or patterns in the data emerged. A second round of open coding ensured that interviews analyzed early and late in the first round were subjected to the same set of codes. Through axial coding I identified relationships among the themes, disaggregating some into finer-grained themes and aggregating others into broader concepts. Finally, through selective coding, I developed a narrative that supports the reporting of findings in ways that are consistent with the research questions. Member checking the findings lends trustworthiness, and I therefore sent or discussed department-specific findings with each admissions chair to refine my conclusions.

Limitations

Findings need to be interpreted in the context of self-selection into the study. Prospective participants gave a range of reasons for nonparticipation, but correspondence suggests that several chairs whom I initially contacted opted not to participate in the study beyond our informational interview out of concern that I would expose unfair practices. Others participated specifically as an opportunity to gain an outside perspective of their admissions process and to learn how they could improve on current practice. Another limitation

is the limited racial diversity in the sample. Sampling the current admissions committees means the data faithfully capture present dynamics of academic selection in these elite programs but that it does not include the voices of many domestic scholars of color. Findings should therefore be read as representing the current state of affairs, not as a model to which programs seeking to improve diversity should aspire.

Findings

Conceptions of merit and diversity are as numerous as the participants themselves, but there are clear patterns in the data that clarify contradictions in the literature and shed new light on the challenges of achieving diversity through holistic review. Perhaps the clearest trends are that conceptualizations of merit vary between initial and latter rounds of review, with different implications for diversity (see table 2). Therefore, in the sections that follow, I separately consider conceptualizations and evaluative scripts of merit and diversity in each round of review.

The First Cut: Quantifying Conventional Achievement

Across programs the initial conceptualization of merit rested on numerical indicators of conventional academic achievement that can be used to quickly compare students. They included GRE and TOEFL scores and grade-point averages, which faculty often contextualized by institutional prestige and curricular rigor. An astrophysicist admitted, “I would say—and you will see it in our discussions—it’s very unlikely that we would consider anyone who has a low subject GRE.” One of his colleagues concurred: “If you don’t score high, you’re probably not going to make the cut.” In physics and astrophysics, scores on the Physics Subject GRE were most important, while the analytical writing score commanded considerable attention from philosophers, linguists, and classicists. One social sciences admissions chair single-handedly winnows the 800 applications received to a still-daunting 400 that a committee reads. About this initial screening, he admitted to me, “Personal statements have almost no role. I don’t read them. I look at the transcript and glance at the GRE. If the quantitative score is not perfect, don’t bother applying.”¹¹

Findings in these 10 programs thus corroborate Klitgaard’s (1985) conclusion that test scores and grades form the “backbone of the evaluative process,” with test scores especially attractive due to their “magic simplicity” (32). Yet, regarding these metrics as a proxy for merit has serious consequences for diversity. Nationally, students of color, women, and low-income students have

TABLE 2

Summary of Evaluation in Initial and Latter Phases of Admissions Review

Phase of Review	
	Short List
<p>Conceptualization of merit Important criteria</p>	<p style="text-align: center;">Full Pool</p> <p>Conventional achievers</p> <ul style="list-style-type: none"> • Test scores • Grade, in context of • Undergraduate institutional prestige • Curriculum rigor <p style="text-align: center;">Short List</p> <p>Future of the discipline</p> <ul style="list-style-type: none"> • Scholarly dispositions • Potential for contributions • Program • Discipline
<p>Bottom-line questions faculty ask</p>	<p>Can the student succeed in our program? Does he seem to present a risk?</p> <p>What will she add? Do his interests fit our expertise?</p>
<p>Relationship of merit and diversity</p>	<p>Standard of merit may be in tension with diversity aims. Diversity is a component of merit.</p>
<p>Evaluative scripts</p>	<p>Scripts of merit:</p> <ul style="list-style-type: none"> • Appropriateness • Consequences • Convenience <p>Scripts of diversity:</p> <ul style="list-style-type: none"> • Obligation • Competition • Opportunity

lower mean GRE scores (ETS 2001; Miller 2013), and there is evidence of grade inflation in highly selective universities where these populations remain underrepresented (Bastedo and Jacquette 2011; Bielby et al. forthcoming; Posselt et al. 2012; Rojstaczer and Healy 2012); therefore, an initial threshold of very high GRE scores and/or grades from selective undergraduate colleges may disproportionately exclude students from already underrepresented groups. Why, then, would faculty continue to rely upon this standard?

Understanding Reliance on Conventional Achievement

Quantitative criteria pass as human capital credentials and have cultural meanings in elite departments, which help explain their enduring influence. By eliciting the meanings of criteria through interviews and triangulating them with observational data, I identified three evaluative scripts for requiring a high quantitative threshold of conventional achievement—each rooted in the programs' prestige within the discipline. *Risk aversion* and *belonging* broadly correspond to March's (1994) logics of appropriateness and consequences, but I also found an important script about the *convenience* of using numbers for initial sifting of the pool. To be clear, not every participant employed each of these justifications; rather, these were the three that faculty most frequently cited.

Risk aversion.—One evaluative script for requiring high GRE scores and grades concerns the imperative and luxury of risk aversion. Concerned about the financial and status consequences of student attrition, faculty viewed their applicants in relation to course requirements, desiring students who were hard workers but who would not struggle with a curriculum that demanded prerequisite skills for which there were few to no remedial learning opportunities. As a classicist put it, "Graduate admissions is one of the things I think you have to be very humble about. And there's always a tension here because we're always under pressure to have good numbers for completing a program, completing it in a reasonable amount of time, and so on. The effect of that is to make you risk averse because it's not that hard just to go for the students you're pretty confident can get through."

A physicist presented his case for risk aversion in terms of time investment and teaching energy:

Interviewer: You've mentioned a couple of times: "Can they be successful?" and "Do I think they'll be successful?" It sounds like that's a really key question you're asking.

Respondent: Yes, because it's a big investment for the faculty member who takes on a student. And so if you work with the student so closely and then he walks away or doesn't make it, then it's a

Toward Inclusive Excellence in Graduate Education

waste of his time and in a way, I mean, it's our mission to teach, but I'd rather spend my time teaching somebody who actually can continue my mission and then teach other students than somebody who realizes, "It's just too difficult. I can't do it."

Across fields, many perceived that undergraduate training and test scores were the best predictors of academic success. They interpreted a weaker record of upper-level course work in the discipline as underpreparation and, as specifically cited in a majority of the programs, a concern the applicant "may not know what she's getting into." Particularly when cohorts were small and/or programs provided full financial support, each enrolled student represented a significant financial investment, so admissions was compared to "risk-taking," "risk aversion," and "gambling."

Participants expressed four different approaches to risk. First, several respondents described the tendency to "feel spooked" when reading the application of a person whose profile reminded them of an unsuccessful student.¹² Second, consistent with a logic of consequences, successful organizations were more likely to attribute their members' outcomes to motivation and skill rather than conditions in the organizational environment, and they therefore tried to eliminate risk rather than estimate it (March 1994). Others, due to credential inflation, conflated variation in the applicant pool with risk of attrition. For example, one member of a philosophy program worried that her colleagues had "a ridiculously high standard" and had become too risk averse, expecting their students to be preprofessional philosophers because typical applicant profiles have tended in that direction. Finally, in an interesting integration of the logics of appropriateness and consequences, four admissions chairs discussed, without my asking, the "luxury of risk aversion" that their large, highly qualified pools permit. Some preferred conventional overachievers simply because they could.

To the extent that admitting any given student might be a risk, it was not a calculated risk, for just one program conducted empirical analysis of applicant qualifications in relation to academic or professional outcomes. But as sociology's Thomas theorem goes, "If men perceive a situation as real, it is real in their consequences." In this case, perceived risks associated with admitting students with weaker "numbers" had consequences for the diversity of the applicant pool in subsequent rounds of review.

Belonging.—As this point suggests, identifying as intellectual communities near the top of their respective fields constrained what participants felt were appropriate selection criteria. They used a script that associated "the numbers" with intelligence,¹³ and intelligence with *belonging* in their community. This script of belonging also helps explain quantitative metrics' enduring influence. In my interviews, more than 50% volunteered comments about GRE scores signaling innate ability (e.g., "sheer intellectual horsepower," "native intelligence"). In admissions meetings, references to general impressions of intelli-

gence were associated more than half of the time with high grades and/or GRE scores. In linguistics, one professor commented, “Those are astronomical scores!” with another responding, “And check out the stellar grades. There’s no question she’s smart.” When I followed up with the department chair, he reiterated, “Someone who does that well on the GRE is unlikely to be lame-brained. They are likely to be smart.”

Research by Campbell (2009) on in-group norms affecting gatekeeping in psychology supports Kierkegaard’s claim that “We create ourselves by our choices.” With intelligence central to both academic culture and the self-concepts of professors (Bourdieu 1988; Lamont 2009), there is a personal quality to these supposedly objective criteria. The legitimacy of departments’ identities as intellectual communities hinges on the perceived intelligence of those whom they admit and hire. As the linguistics committee chair put it, they used admissions to “reflect the view the department has of itself” because the department’s character is “so determined by graduate admissions.” Linking quantitative criteria with intelligence and belonging explains reliance on quantitative criteria, but belonging mattered throughout the review process, for “fit” was an almost unassailable criterion and, often, a bottom-line consideration. For example, explaining rejection of a borderline applicant, a biologist commented, “[He was] from a different planet and we were confident that this person was not going to be one of us. He’s not going to be a full member of the scientific community.”

Recall that in March’s (1994) logic of appropriateness, an organization makes decisions by assessing the situation, its identity, and what such an organization should do in such a situation. In admissions, some faculty advanced to full committee review those with the highest GRE scores and grades because they thought of themselves—tacitly or explicitly—as an elite intellectual community, of admissions as a means of building that community, and of high standards of conventional achievement as consistent with that identity.

Convenience.—A third evaluative script for initially relying on metrics of conventional achievement is not related to logics of appropriateness or consequences, but rather the convenience of comparing applicants through “the numbers.” Time demands and incomplete information were the most frequently cited answers to my interview question about what makes admissions evaluations difficult. And although ETS recommends against the practice, some committees took advantage of the apparent comparability of student test scores to sort long lists of applicants by scores and grades with the click of a mouse on a spreadsheet. One committee chair in the physical sciences requested, “In reports to me, just summarize the test scores and GPA because I’m fully capable of reading a spreadsheet, but I prefer not to have to read the entire file.” In larger programs, some expressed regret that their colleagues did not give every file a full, detailed reading in the first round of evaluation.

Toward Inclusive Excellence in Graduate Education

A sociologist explained: “We receive so many applications, and we are always in a crunch with time—always. And I have impressions that some of my faculty—senior members—were simply looking at the GRE. They have a threshold such as, ‘If it’s not over 700, I won’t read anything.’ And that cuts usually two-thirds of the applicants.”

Cultures of gatekeeping in these programs were not just about judgments of quality; they wanted to judge quality quickly and efficiently. To interpret letters of recommendation, participants claimed one must “read between the lines” and “sift through the superlatives.” Personal statements were thought of as “subject to gaming,” and they felt it was not only time consuming to read writing samples but also difficult to determine the extent of faculty involvement in crafting the ideas or narrative. On the other hand, faced with hundreds of applications, numbers had an apparent—if illusory—clarity, simplicity, and precision. Despite doubts about their validity, test scores and grades seemed to distill constructs as complex as ability, preparation, and achievement into a single number on which everyone can be quickly compared. A sociologist explained: “The GRE is something they all have in common. The fact that it is common to everybody is really useful.” Valuing scores and grades, then, was not only a matter of risk aversion but also of ambiguity aversion. In the absence of clarity, faculty settle for the convenience of the entrenched standard.

To summarize, faculty in these programs most frequently used scripts of appropriateness, consequences, and convenience to justify a conventional, quantitative conceptualization of merit in making their first cut. With underrepresented students less likely to earn the high scores that propel an application to the next round, rigid adherence to this notion of merit may have undermined institutional diversity aims. However, applications that could surmount the high initial threshold were subjected to true holistic review and were assessed against a different conceptualization of merit.

Rethinking Merit for the Short List: The Future of the Discipline

With the richer information introduced in qualitative elements of the application, faculty employed a broader conceptualization of merit in later rounds of evaluation. It was not enough at this point for applicants to have mastered course material and standardized tests or earned the favor of scholars in top colleges and universities, for these programs had large numbers of such conventional achievers. Reviewers were looking for applicants with a different relationship with knowledge. Applicants’ apparent potential for original scholarship, including innovation born of fresh perspective, became central to perceptions of admissibility.

In holistically reviewing the files on their short lists, faculty tried to divine

which individuals could grow to become leaders in their respective fields. Respondents from disciplines that were far afield intellectually expressed similar perspectives on the difficulty of this predictive work. An astrophysicist said: “You’re really basing it on their potential to be great scientists. But it is much harder—much harder. What you look at for a particular person varies from person to person. Some people just haven’t had a good research experience and you’re basing it on more numerical factors or letters from people you don’t know.”

And an economist elaborated:

That’s so hard because at one level what you’re looking for is so easy. You’re basically looking for people who are going to be first-rate researchers and leaders in the field . . . who are going to be great economists. But figuring out in a 21-year-old what traits are predictive of becoming great intellectual leaders is incredibly hard. . . . You want to find someone who is creative and asks great questions. That is so hard to tell. You might look at their thesis or maybe they worked as a research assistant or their recommendations. It’s very hard to assess if someone is—*[pause]* it takes a certain type of person to be a researcher.

In classics, where language training and travel experience were as important as mathematical training in economics or research experience in astrophysics, professors also struggled to identify talent:

From the department’s perspective, we want to have somebody who is going to finish the program and thrive. Thrive here and go on and be a successful faculty member somewhere. So we want to be careful about who we pick. . . . They’re really pretty young, just finishing college and a lot of students really still are quite unformed in a lot of ways, still taking a lot of other classes on top of their class courses. What will that person be like in 1 or 2 years? You know, where will they go? That’s where the guesswork comes in. I think the more experienced—the more you teach, the more you know students, the more you choose them and see how they turn out, the better feeling you have for it. But there’s something. It’s not quantity. A score doesn’t tell you, answer that question.

Across the disciplines, faculty valued displays of potential to become what Walker et al. (2007) call a steward of the discipline, one who “considers the applications, uses, and purposes of the field and favors wise and responsible application . . . [and] how to foster renewal and creativity” (11). Thus, participants assessed not only human capital but also personal characteristics that condition how human capital is deployed, including affective dispositions and noncognitive traits that they associated with exemplary scholarship. A physical

Toward Inclusive Excellence in Graduate Education

scientist explained (emphasis mine): “I look for people who have had research *experience* and show a great deal of *enthusiasm* for it. I look for *creativity*. I look for *energy*. I look for possibilities, potential for *innovation* . . . a *passion* for research.”

And in a social sciences program: “If the person is intellectually *creative*, I think he or she *will grow to be a really, good scholar*. To me that is the most important trait I look for. . . . He or she is *really thinking*. It’s not like—I know this language. I have that skill. I have learned these courses, all this typical knowledge. I’m *curious* about all of the new, interesting phenomena in society or this or that. And I *want to explore*. I want to find what’s going on. . . . That would make a good scholar.”

Such characteristics as curiosity, passion, and creativity connote a relationship to knowledge that cannot be measured quantitatively but that may be inferred from elements of the application that many claimed were too onerous to read carefully in the first round. Faculty tried to glean these qualities through student or recommender descriptions of research experience so they could evaluate potential as a function of prior accomplishment, as in hiring faculty: “I wanted to see some independent research because I think if you haven’t done independent research, then you don’t really know whether you have a passion for it. And you kind of don’t know what you’re getting yourself into. This is what it is all about.”

And in political science and astrophysics, reviewers placed a special premium on applicants’ ability to “sell your ideas.” When pro-research dispositions, scholarly potential, and fit with department expertise converge in an applicant, faculty responded favorably. “We don’t want clones,” one humanities professor laughed. “But on the other hand, we do want people who really do match reasonably well to what we’re good at.”

Evaluative Scripts of Diversity

In later phases of review, faculty talked a lot about diversity, and it was an important part of how they thought about merit. Invariably, they displayed preferences for applicants who would add to the program’s and/or discipline’s diversity. In the presence of widely divergent ideas about how diversity should be defined, there was a clear pattern to their understandings of why diversity is important. In this section, I discuss the evaluative scripts of *obligation*, *opportunity*, and *competition* that motivated how committees deliberated and justified their consideration of diversity.

Obligation.—First, many faculty associated diversity with social obligation, and they approached admissions with a base sense of obligation to improve the representation of women and underrepresented racial minorities in their

programs. However, they also felt obligated to factor diversity into their ratings in ways that protected the entrenched standard of conventional achievement. A sociologist of color summed up this perspective: “You have to be above a bar, and then we can ask the diversity question.” One of his colleagues put it this way:

That’s one of the difficult issues, because I would say this: I try not to pay too much attention. I try to admit students that are the best in my intellect with no regards for gender or race. And so I understand, I perfectly understand, that because the world out there is not equal, this simply reproduces inequalities. I understand that. But at the same time—so I will say this—if there are two students that are, in my view, equal on intellectual merit, then I will prefer a minority. And I think it is the same—at least from what I observed—it probably was the same for my other colleagues on the committee. It’s something that everybody pays attention to—the minority race, gender—for good reason.

According to professors in both philosophy programs, field-wide attention to women’s underrepresentation in philosophy has heightened the sense of obligation to admit more women. Awareness that other “macho professions,” as a few philosophers called theirs, seemed to be making better progress only sharpened their sense of duty to improve the pipeline. A male philosopher elaborated: “We were sort of going backwards while the fields that you think of as the most paradigmatically male fields were inching toward something.” Philosophy was also the only field in which I observed debate about when diversity should be considered; however, the one committee that did debate this eventually landed on the same practice as other programs. To achieve “diversity with excellence,” their practice was to “select out the top students and seek in the second round to weight the students who meet diversity criteria.” In sum, most felt obligations to diversity were contingent on maintaining a standard of “excellence” that few seem willing to revise.

Competition.—One reason reviewers felt obligated to consider diversity conditional on a conventional approach to merit was because other top programs in the discipline applied the same standard. Attracting academically accomplished students from underrepresented backgrounds has become a way that programs evaluate themselves against one another, such that diversity itself is associated with prestige. As a result, when the traditional standard of merit was fulfilled, applicant contributions to diversity were highly valued, especially if the student identified as an underrepresented gender or race/ethnicity. Of an underrepresented minority student who attended an Ivy League institution, a philosophy professor made the offhanded comment, “Sounds like he’ll get in everywhere. Everyone will love him.” Faculty in three programs agonized over their yield in both interviews and observations, viewing the ability to

Toward Inclusive Excellence in Graduate Education

attract top students as a hallmark of the program's positioning relative to others. The ability of a program to successfully entice an underrepresented student away from a peer institution both was thought to further its own diversity and to affirm its legitimacy.

Therefore, just as students with high test scores have been for decades, high-achieving students of color and, in some fields, women, have become objectified as a highly valued organizational asset. Admissions deliberations included repeated comments such as, "Can we get her?" "Who are we going to get? It's a gamble," and "We'll lose him to Princeton and Caltech." I inquired into these offhanded comments in follow-up interviews. Participants expressed a belief that underrepresentation is not only a function of who was in the initial pool of applicants or the ways in which some criteria filtered out promising students of color but also that talented applicants of color tended to have many admissions offers. In economics, a professor discussed the value of intellectual diversity in relation to their struggles to attract black students:

I think people care most about intellectual diversity—that people arrive with different interests, different preparations, and are likely to write very different theses on different topics. I think that's the type of diversity we would value most. I mean gender is an issue in that we get good—we get top-notch women as well top-notch men. Black—we get fewer blacks. It's true. But we do try—in the past we've tried to attract them. But then they get the same attractive offers from Columbia and Yale and Stanford and Berkeley and so forth. So it's a small group typically who get a lot of attention. We look at a big pool of the world.

Thus, even in top programs, faculty felt that one challenge in achieving diversity lies not with their choices of applicants but students' choices among their offers. A sociologist of color notes: "We all kind of admit the same pool of applicants—the top 10 departments. Harvard's going to admit them, Princeton's going to admit them, Stanford, Columbia, Michigan, Wisconsin, and Chicago. And so we're all fighting for the same applicants, and there's a lot to compete with, and you know there are a couple programs right now that are just kind of doing phenomenally well in terms of placement and training—just dominating sociology."

However, such claims must be interpreted in the context of weak recruitment efforts in some programs and weak coordination of recruitment and admission in others. Stepping up recruitment before applications are submitted would surely help, but some were hesitant to do so because they felt it would introduce diversity considerations into the process earlier than they feel comfortable.

Opportunity.—Participants saw diversity's desirability primarily in terms of the benefits it has for the program. In addition to prestige, as discussed above,

another set of organizational benefits concerned the intellectual and financial opportunity for the program they felt diversity offered.

Intellectual opportunity: Websites for all three universities' graduate schools discussed diversity's centrality to the mission of graduate education and ideas of excellence. At the department level, some faculty associated diversity with intellectual opportunity. Responding to the question about how their department conceptualizes diversity, a professor of classics said: "We welcome it. We want it. It's so much. I think everybody is committed to it on principle. And we know from past experience how much it enhances our classrooms and our life in our department. And the university really supports it, I think. . . . Students who come from other backgrounds will have other—will just focus on other, um, moments or ideas in the text. Also, students who have different backgrounds will have studied different things. There's a kind of desire, there's [*pause*] more of a drive in some of the students who come from different backgrounds."

And, indeed, diversity considerations often went deeper than whether a student fits into a specified racial, gender, or other social category. Faculty compared students within categories against one another, reading into information conveyed in the personal statement to discern differences in identity-related perspectives that two students from the same demographic category might offer.

A deliberation in astrophysics exemplified this tendency, although I also observed similar conversations in classics, philosophy, linguistics, and political science. Discussing one female applicant, a professor posed the question, "Is it enough to be a woman in science?" They discussed how women in the pool had used the personal statement to disclose gender-related experiences in science and the sort of experiences they thought might positively contribute to the community they were trying to build. For example, they noted that one applicant wanted to be a role model for younger women because she had never personally received explicit encouragement. Another expressed a need to develop self-confidence and overcome self-doubt.

The committee went on to comparatively evaluate these applicants' narratives against one another, speculating what the implications might be of having such individuals in the department. In reference to one who said she faced "teasing and bigotry" from her teachers and peers, a male committee member remarked, "I'm less persuaded by that story," defending the teacher and suggesting the student might come to the department with an axe to grind. His opinion was overruled, however, after another male committee member noted, "Now she's taking action, organizing a lecture series" on women in science.

This example demonstrated the tendency for identities and experiences of so-called diversity candidates to be scrutinized at a level that applicants from majority backgrounds were not. This finding is consistent with prior research.

Toward Inclusive Excellence in Graduate Education

Eighty percent of subjects judging a Latino and 75% of those judging a white female for positions in higher education administration cited the doctoral institution as very important, but only 55% of those evaluating a white male cited it as very important (Haro 1995). Similarly, Danowitz-Sagaria (2002) found equal outcomes of selection across race and gender, but she found that African American women applying for administrative positions in universities were subjected to “filters” (i.e., sets of criteria) that white men and women were not.

Social categories were just the start of understanding what applicants might bring to a program in the name of diversity. Committee members sometimes compared students in underrepresented categories with one another, interpreting disclosures about identity as opportunities or red flags in ways that ultimately affect selection outcomes. The conversations I observed conformed to what Chang (2002) calls a diversity discourse of preservation (vs. one of transformation), in that participants sought diversity as an intellectual opportunity that would not disrupt the status quo. The institutional-level message conveyed in websites and publications may have been one of unequivocal support for the intellectual opportunity that diversity represents, but at the level of file review faculty regarded some applicants contributing diversity that was more appealing than others.

Financial opportunity: Faculty also considered financial opportunity in relation to student diversity. All three universities, under statements about diversity’s role in enhancing the institution, made available a number of 4-year, incremental, and matching fellowships to programs that admitted students who contribute to diversity by race/ethnicity, gender, and social economic status. Faculty reiterated over and over again what a powerful incentive this was. In some cases, committees adopted the university’s conceptualization of diversity to ensure they were maximizing the opportunity. On asking an associate professor of sociology how the department conceptualizes diversity, he admitted:

I think roughly in whatever way the university will pay for. Our conceptualization is the university’s conceptualization, and that’s putting it a little harsh. But because the university’s commitment is quite good and there are lots of incentives, we don’t need to add any interests that are beyond what are well established through straightforward incentives. So we’ll just do whatever. We define diversity as the university defines it. If this person’s going to be eligible for some sort of special resources, then we’re all for it. I don’t see much for it beyond the university’s commitment as implemented through a variety of programs that give us resources we need to be more diverse.

Although he admitted he was putting it “a little harsh,” the fellowships clearly offered a strong incentive.

Taking advantage of diversity fellowships involved a subselection process within the broader admissions process, including deliberation about which applicants should be nominated and what counts as diversity. In astrophysics, the committee chair encouraged his colleagues to very carefully read applications for evidence of both background factors and contributions to diversity. “We need to read between the lines on these things,” he said, to nodding around the room. Once discussion began, the comments and the type of advocacy I heard suggested a strong interest in nominating underrepresented racial minorities. The committee walked a fine line, wanting to be inclusive in their conception of diversity, yet caring very much about the serious issues of underrepresentation in their field. They worried in particular about fellowship criteria that may map onto race/ethnicity (e.g., service, coming from an urban college) instead of selecting on race directly, because they might miss some who would have qualified as “minorities” while picking up others who would not help remediate stratification. This rich discussion ended without clear answers, but it bears mentioning that the committee members never explicitly mentioned the race or ethnicity of the student whom they ultimately nominated.

To summarize, when faculty evaluated students who had made it to the short list, they invariably considered both qualitative and quantitative elements of the application. Substantively, they privileged students who seemed to have the research experience and noncognitive dispositions to become leaders and innovators in the field. At this stage, diversity was central to ideas about merit, and the faculty justified it through evaluative scripts about obligation, competition, and intellectual and financial opportunity. Sometimes principled, sometimes pragmatic, these norms served a wide range of faculty interests. Yet by waiting to weigh diversity until reviewing the short list, the criteria associated with conventional achievement carried the day in shaping access to these selective doctoral programs. Diversity may have mattered a great deal for individual borderline cases, but it did not reduce the prominence of high GRE scores and grades from elite colleges in the process.

Discussion

In this article, I have examined how faculty in highly ranked programs constructed merit and considered diversity in doctoral admissions. I have explored diversity considerations in relation to merit, and I have distinguished the conventional achievers programs initially identify using quantitative metrics from the future leaders of the discipline reviewers sought to identify through holistic review. I have also outlined evaluative scripts, or decision-making pathways, by which faculty give meaning to their ideas about specific criteria.

Toward Inclusive Excellence in Graduate Education

Findings support previous research about the importance of GRE scores, undergraduate grades and institutional prestige, race/ethnicity, and perceptions of professional competency to judgments of admissibility (Attiyeh and Attiyeh 1997; Brear et al. 2008; Campbell 2009; Eide et al. 1998; Gartner 1986). Qualitative research also enabled me to uncover why faculty privilege these criteria.

Although many participants expressed distrust in or disdain for the reliability of quantitative metrics like test scores and grades, these metrics were inseparable from ideas about merit early in the review process under scripts of convenience, belonging, and risk aversion. More than half assumed “the numbers” signaled intelligence, which they associated with belonging in their community and the student’s likelihood of success. At least as important, however, these measures of conventional achievement were tools of convenience in efforts to narrow the applicant pool to a size for which holistic evaluation seemed manageable. If applicants surmounted the informal numeric thresholds required in early phases of review, their applications were subjected to an entirely different, more holistic, reading in later rounds. Rather than assessment by numerical proxy, faculty interacted with a person behind the application and the numbers. Here, they weighed potential for innovation in the discipline, fit of their research interests with departmental strengths, and ways (including through diversity) that they would contribute to the future of the department and discipline.

In both phases, faculty saw graduate students as key members of their intellectual community, so their boundaries in selecting students reflected a desire for those who “fit.” These are culturally embedded preferences, or tastes, defined not only by organizational power plays as would be expected according to Bourdieu’s (1984) theory of taste and selection but also by doctoral education’s mission of collective knowledge advancement and stewardship (Walker et al. 2007).

Through a qualitative analysis of the entire cycle of admissions evaluation, this research helps inform the emerging literature about the role of diversity in allocating academic opportunities. Grodsky (2007) found that unequal distributions of SAT/ACT test scores inhibited undergraduate admissions opportunities for students of color. On the other hand, Lamont (2009) reported that diversity was a strong preference in selection for interdisciplinary post-doctoral fellowships. Evidence in the present study helps reconcile these views. In the later, more holistic rounds of admissions review, diversity contributions indeed surfaced as an important value, and participants felt socially obligated to improve enrollment of women and students of color. However, they also felt obligated not to consider diversity from the outset of the review process, but rather to focus on quantitative metrics that advantage overrepresented

populations. As such, many of the students whose diversity contributions might have been considered assets had already been filtered from the pool.

The current two-tier review process relegates diversity to a secondary consideration, and it makes a standard of inclusive excellence conditional on conventional achievement. None of the 10 programs began with diversity as a major criterion. Yet, if a program deeply values diversity, then when they value it matters for the outcomes they are likely to achieve.

Moreover, that one of the clear evaluative scripts of diversity involves status competition corroborates Stevens's (2008) conclusion that "'Diversity,' as measured by the number of students in sharply defined categories is now an index of prestige" (182). I found that faculty sought applicants from specific categories, but not in an absolute sense. Rather, they comparatively evaluated students in these categories against one another, making value judgments about the relative worth of possible diversity contributions. As exemplified in discussion of female applicants in astrophysics, faculty did not reflexively give preference to individuals from specific categories. They dug into the personal statements and letters of recommendation for evidence that the student would bring unique, positive, resilient perspective with their membership in particular groups.

The theoretical framework proposed to understand faculty evaluation in admissions as professional gatekeeping and organizational boundary maintenance. The outcomes of admissions evaluation condition enrollment, which conditions degree attainment and labor market transitions. As such, diversity in graduate education conditions the possibilities of diversity in academia and a range of professions. Levels of underrepresentation in graduate school enrollment and doctoral degree conferral are similar to those observed among faculty (Aud et al. 2010). African Americans and Latinos/as comprise 12% and 15% of the US population, for example, but only 3% of US life and physical scientists (National Science Foundation 2004; US Census Bureau 2008). The problem is evident in the humanities as well, where only 21% of employed philosophers are female (Crasnow 2007) and Asian Americans are critically underrepresented (Altbach et al. 2002). Structurally, realizing inclusive excellence in graduate education is prerequisite to realizing it in the professoriate and other professions.

Future Research

There is much we have yet to learn about graduate admissions. This study begins a conversation about faculty evaluation as a matter of advanced educational access, articulating processes of evaluation, the multivalent nature of merit that faculty construct, and rationales for those constructions. Future research needs to consider not only the barriers to equity and diversity in

Toward Inclusive Excellence in Graduate Education

graduate education but also what works. How do cultures of graduate programs change over time? What can we learn from programs that are both selective and diverse? How do they resist field-level pressures for conventional achievers? Research should also examine practical efforts, such as faculty-administrative coordination of admissions with recruitment and the effects of institutional diversity fellowships on admission, enrollment, and degree completion. We also need to know more about self-selection into graduate school, as well as how growing demand from international applicants complicates notions of racial and ethnic diversity in US graduate education.

Conclusion

Recall that scripts can be thought of as the stories we tell ourselves and others to justify routine behavior. In this case, many faculty claimed a desire for equity, and they acknowledged diversity's organizational and educational benefits, but they felt it was inappropriate to give up an extraordinarily high standard of academic and test score performance given their prestige orientations. They also worried about the consequences of broadening this standard given its perceived associations with later success, and they argued that it was simply inconvenient given the volume of applications they receive and the ambiguity of holistic, comparative evaluation.

By framing these justifications as constraints, however, they failed to acknowledge that these structures and filters reflected choices they had made for their program—and that they continued to tacitly make through the inertia born of unexamined practice. They could structure their required course work differently, for example, and empirically assess the risk hypothesis by analyzing which admissions characteristics are associated with success in their program or discipline.

Conventional achievement was the standard against which faculty assessed all applicants, but the evidence made clear that it was not the only set of considerations. That they assessed the shortlist by considering stewardship of the discipline—including research dispositions, noncognitive strengths, and diversity contributions—indicated that multiple hierarchies of value, or heterarchies, guided evaluations in admission. And that even in the most exclusive program in this study, the chair has begun to contextualize black and Latino students' GRE scores indicates that resistance to considering diversity early in the review process may be softening.

Further, evaluations of applicant qualities were flexible to other organizational interests, such as faculty collegiality and preserving a balance of students across department concentrations. Some meanings of diversity were principled, such as regarding it as a social obligation or opportunity to enrich the de-

partment's intellectual climate. However, consistent with Boltanski and Thévenot's (2006) discussion of pragmatic compromise over heterarchies, participants in this study deferred in many cases of disagreement to pragmatic organizational interests over personal principles. Scripts for valuing diversity illustrated this pragmatism, especially the tendency to associate diversity with financial opportunity and competition with other highly ranked programs in the discipline.

Juxtaposing scripts of merit and diversity produces a list of values and premises that motivated decision making and which scholars used to rationalize current practice in these highly selective programs. Belonging, risk aversion, convenience, obligation, competition, and opportunity can be thought of as part of a bounded rationality for members of these academic communities. Change at the program level will require organizational learning that has both cultural and structural dimensions. Structurally, reforms to graduate admission can work with the current pragmatism rather than against it by developing more efficient approaches to holistic review and strengthening incentives for diversity. And, culturally, present values need to be interrogated, because they evince an organizational culture in which notions of quality are caught up first with prestige, and only secondarily with principles of equity and diversity. Moving toward a culture of inclusive excellence will require counterscripts—fresh understandings about what admissions considerations mean—and collective engagement by faculty and administrators as arbiters of educational opportunity.

Notes

1. I have masked and/or changed all information that might be personally identifiable. In referring to colleges and universities, I sought to balance the need to ensure anonymity with the need to convey a real-world sense of the institutional strata in which these programs are located. Therefore, when naming specific universities including the data collection sites, I randomly drew from a list of 15 universities in the same tier of program rankings for that discipline. This means that since the actual university in which data were collected could be named in the article due to chance, readers should be no more able to recognize it as being the data collection site than any other similar institution.

2. Research on graduate admissions initiated in the mid-twentieth century to analyze the validity of standardized tests and other graduate admissions criteria in predicting admitted students' success (e.g., Cureton et al. 1949; Marston 1971; Newman 1968; Pfeifer and Sedlacek 1971).

3. Just one study has specifically probed the effects of diversity on graduate students' educational experience, in medical education (Whitla et al. 2003). However, faculty instruction of undergraduates can reinforce broader institutional diversity goals (Mayhew et al. 2005), a relationship that may extend to graduate students given the closer relationships they tend to have with faculty.

4. In this article, I examine trends that apply across disciplines; however, disciplinary

Toward Inclusive Excellence in Graduate Education

dynamics are also salient in graduate admissions and are the focus of forthcoming research from this study.

5. Bourdieu and Passeron (1977) name as elites those with high levels of occupational prestige, years of schooling, and income.

6. Although these ideas are not central to my analysis, Bourdieu's own definitions of these key concepts, and their relation to one another and to habitus and field, are most succinctly laid out in "The Forms of Capital" (Bourdieu 1986).

7. Evaluative scripts are distinct from interaction scripts qua Barley (1986): "outlines of recurrent patterns of interaction that define in observable and behavioral terms, the essence of actors' roles" (83).

8. Although PhD *programs* are nested within academic *departments*, many participants used these words interchangeably. I use program to describe the context of my data collection.

9. Considering that the sample largely represents the current constitution of admissions committees in each program, these numbers provide additional evidence of the need to improve gender and racial diversity in highly ranked PhD programs.

10. I observed admissions committee meetings in astronomy, classics, linguistics, both philosophy programs, and political science.

11. In the course of member checking, however, he made a point of mentioning that he now pulls applications from black and Latino students to ensure a diverse pool in later rounds of review and that they have relaxed the GRE score threshold since the test's 2011 revision, which "nicely distinguishes among highly qualified people."

12. Judging a person by their impressions of another with some similar traits is a cognitive heuristic and a natural propensity, but it is an attribution bias, especially when making the comparison on the basis of a small sample of cases.

13. The ETS does not claim that the GRE measures intelligence. The company has worked hard to refine the instrument away from the format of IQ tests and to provide training and written materials to instruct on what the test measures and how scores should be interpreted and used. For guidance on interpreting scores on tests taken after August 2011, see <http://www.ets.org/gre/institutions/scores/interpret>.

References

- Altbach, Philip, Kofi Lomotey, and Shariba Rivers. 2002. "Race in Higher Education." In *The Racial Crisis in American Higher Education*, ed. Philip Altbach and Kofi Lomotey. Albany: SUNY Press.
- Antonio, Anthony. 2004. "When Does Race Matter in College Friendships? Exploring Men's Diverse and Homogeneous Friendship Groups." *Review of Higher Education* 27 (4): 553–75.
- Association of American Colleges and Universities. 2013. "Making Excellence Inclusive," http://www.aacu.org/compass/inclusive_excellence.cfm.
- Attiyeh, Gregory, and Richard Attiyeh. 1997. "Testing for Bias in Graduate School Admissions." *Journal of Human Resources* 32 (3): 524–48.
- Aud, Susan, Mary Ann Fox, and Angelina Kewal-Ramani. 2010. "Status and Trends in the Education of Racial and Ethnic Groups" (NCES 2010-015). US Department of Education, National Center for Education Statistics, <http://nces.ed.gov/pubs2010/2010015.pdf>.
- Azoulay, Pierre, Christopher Liu, and Toby Stuart. 2009. "Social Influence Given

- (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs." Unpublished manuscript, Harvard Business School, Cambridge, MA.
- Barley, Stephen R. "Technology as an Occasion for Restructuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments." *Administrative Science Quarterly* 31 (1): 78–108.
- Bastedo, Michael N., and Ozan Jaquette. 2011. "Running in Place: Low-Income Students and the Dynamics of Higher Education Stratification." *Educational Evaluation and Policy Analysis* 33 (3): 318–39.
- Bielby, Rob, Julie R. Posselt, Ozan Jaquette, and Michael Bastedo. Forthcoming. "Why Are Women Underrepresented in Elite Colleges and Universities? A Nonlinear Decomposition Analysis." *Research in Higher Education*.
- Birnbaum, Robert. 1988. "Presidential Searches and the Discovery of Organizational Goals." *Journal of Higher Education* 59 (5): 489–509.
- Boltanski, Luc, and Laurent Thévenot. 2006. *On Justification: Economies of Worth*. Princeton, NJ: Princeton University Press.
- Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Cambridge, MA: Harvard University Press.
- Bourdieu, Pierre. 1986. "The Forms of Capital." In *Handbook of Theory and Research for the Sociology of Education*, ed. John G. Richardson. New York: Greenwood.
- Bourdieu, Pierre. 1988. *Homo Academicus*. Stanford, CA: Stanford University Press.
- Bourdieu, Pierre, and Jean-Claude Passeron. 1977. *Reproduction in Education, Society and Culture*. Thousand Oaks, CA: Sage.
- Brear, Pamela, Jillian Dorrian, and Giuseppa Luscri. 2008. "Preparing Our Future Counselling Professionals: Gatekeeping and the Implications for Research." *Counselling and Psychotherapy Research* 8 (2): 93–101.
- Brink, William. 1999. "Selecting Graduate Students." *Journal of Higher Education* 70 (5): 517–23.
- Bryson, Bethany. 1996. "'Anything but Heavy Metal': Symbolic Exclusion and Musical Dislikes." *American Sociological Review* 61 (5): 884–99.
- Campbell, Joanna. 2009. "Attitudes and Beliefs of Counselor Educators toward Gatekeeping." PhD diss., Old Dominion University.
- Carter, Dorinda. 2008. "Achievement as Resistance: The Development of a Critical Race Achievement Ideology among Black Achievers." *Harvard Educational Review* 78 (3): 466–97.
- Chang, Mitchell. 2002. "Preservation or Transformation: Where's the Real Educational Discourse on Diversity?" *Review of Higher Education* 25 (2): 125–40.
- Corbin, Juliet, and Anselm Strauss. 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: Sage.
- Crasnow, Sharon. 2007. "Report on the APA Status of Women." Paper presented at the annual meeting of the Central Division American Philosophical Association, Chicago.
- Cureton, Edward E., Louise W. Cureton, and Ruth Bishop. 1949. "Prediction of Success in Graduate Study of Psychology at the University of Tennessee." *American Psychologist* 4:361–62.
- Danowitz-Sagaria, Mary Ann. 2002. "An Exploratory Model of Filtering in Administrative Searches: Toward Counter-Hegemonic Discourses." *Journal of Higher Education*, 73 (6): 677–711.
- Denson, Nida, and Mitchell Chang. 2009. "Racial Diversity Matters: The Impact of Diversity-Related Student Engagement and Institutional Context." *American Educational Research Journal* 46 (2): 322–53.
- Eide, Eric, Dominic Brewer, and Ronald Ehrenberg. 1998. "Does It Pay to Attend

Toward Inclusive Excellence in Graduate Education

- an Elite Private College? Evidence on the Effects of Undergraduate College Quality on Graduate School Attendance.” *Economics of Education Review* 17 (4): 371–76.
- Emerson, Ralph Waldo. 1875. *Letters and Social Aims*. Boston: Houghton Mifflin.
- Espeland, Wendy, and Michael Sauder. 2007. “Rankings and Reactivity: How Public Measures Recreate Social Worlds.” *American Journal of Sociology* 113 (1): 1–40.
- Espino, Michelle. 2014. “Exploring the Role of Community Cultural Wealth in Graduate School Access and Persistence for Mexican American PhDs.” *American Journal of Education* 120 (4) (in this issue).
- ETS (Educational Testing Service). 2001. “Differences in the Gender Gap: Comparisons across Racial and Ethnic Groups in Education and Work.” Policy Information Center, Educational Testing Service, Princeton, NJ, <https://www.ets.org/Media/Research/pdf/PICGENDER.pdf>.
- Fisher v. University of Texas at Austin et al.* 2013. 11-345 U.S. (2013).
- Garces, Liliana. 2011. “Considering Race in Graduate Admissions: Have Statewide Affirmative Action Bans Reduced the Rates at Which Students of Color Enroll in Graduate Programs?” PhD diss., Harvard University.
- Gartner, John. 1986. “Antireligious Prejudice in Admissions to Doctoral Programs in Clinical Psychology.” *Professional Psychology: Research and Practice* 17 (5): 473–75.
- Glaser, Barney, and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Overlook.
- Grodsky, Eric. 2007. “Compensatory Scholarship in Higher Education.” *American Journal of Sociology* 112 (6): 1662–1712.
- Grutter et al. v. Bollinger et al.* 2003. 539 U.S. (2003).
- Guinier, Lani. 2002. “Race, Testing and the Miner’s Canary.” *Rethinking Schools* 16 (4): 13–23.
- Guinier, Lani. 2003. “Admissions Rituals as Political Acts: Guardians at the Gates of Our Democratic Ideals.” *Harvard Law Review* 117 (113).
- Gumport, Patricia. 1993a. “Graduate Education and Organized Research in the United States.” In *The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan*, ed. Burton Clark. Berkeley: University of California Press.
- Gumport, Patricia. 1993b. “Graduate Education and Research Imperatives: Views from American Campuses.” In *The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan*, ed. Burton Clark. Berkeley: University of California Press.
- Gurin, Patricia, Eric Dey, Sylvia Hurtado, and Gerald Gurin. 2002. “Diversity and Higher Education: Theory and Impact on Educational Outcomes.” *Harvard Educational Review* 72 (3): 330–67.
- Gurin, Patricia, Jeffrey Lehman, Earl Lewis, with Eric Dey, Gerald Gurin, and Sylvia Hurtado. 2004. *Defending Diversity: Affirmative Action at the University of Michigan*. Ann Arbor: University of Michigan Press.
- Haro, Roberto. 1995. “Held to a Higher Standard: Latino Executive Selection in Higher Education.” In *The Leaning Ivory Tower: Latino Professors in American Universities*, ed. Walter Bennis. Albany, NY: SUNY Press.
- Hurtado, Sylvia. 2001. “Linking Diversity and Educational Purpose: How Diversity Affects the Classroom Environment and Student Development.” In *Diversity Challenged: Evidence on the Impact of Affirmative Action*, ed. Gary Orfield. Cambridge, MA: Harvard Education Publishing Group.
- Kahlenberg, Richard. 1996. *The Remedy: Class, Race, and Affirmative Action*. New York: Basic.

- Karabel, Jerome. 2005. *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. New York: Houghton Mifflin.
- Karen, David. 1990. "Toward a Political-Organizational Model of Gatekeeping: The Case of Elite Colleges." *Sociology of Education* 63 (4): 227–40.
- Killgore, Leslie. 2009. "Merit and Competition in Selective College Admissions." *Review of Higher Education* 32 (4): 469–88.
- Klitgaard, Robert. 1985. *Choosing Elites: Selecting the "Best and Brightest" at Top Universities and Elsewhere*. New York: Basic Books.
- Lamont, Michele. 1992. *Money, Morals, and Manners: The Culture of the French and American Upper-Middle Class*. Chicago: University of Chicago Press.
- Lamont, Michele. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
- March, James. 1994. *A Primer on Decision Making: How Decisions Happen*. New York: Free Press.
- Marston, Albert R. 1971. "It Is Time to Reconsider the Graduate Record Examination?" *American Psychologist* 26 (7): 653–55.
- Mayhew, Matthew, Heidi Grunwald, and Eric Dey. 2005. "Curriculum Matters: Creating a Positive Climate for Diversity from the Student Perspective." *Research in Higher Education* 46 (4): 389–412.
- Milem, Jeffrey, Mitchell Chang, and Anthony Antonio. 2005. *Making Diversity Work on Campus: A Research-Based Perspective*. Washington, DC: Association of American Colleges and Universities.
- Miles, Matthew, and Michael Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage.
- Miller, Casey. 2013. "Diversity in Physics: Impact of Using Minimum Acceptable GRE Scores for Graduate Admissions." *Bulletin of the American Physical Society*, no. 58, <http://meetings.aps.org/Meeting/MAR13/Session/F38.3>.
- National Center for Institutional Diversity. 2011. "Mission Statement," <http://www.ncid.umich.edu/about/mission.shtml>.
- National Science Foundation. 2004. "Women, Minorities, and Persons with Disabilities in Science and Engineering, 2004." NSF 04-317. National Science Foundation, Division of Science Resource Statistics, Arlington, VA.
- Neumann, Anna. 1991. "The Thinking Team: Toward a Cognitive Model of Administrative Teamwork in Higher Education." *Journal of Higher Education* 62 (5): 485–513.
- Newman, Richard. 1968. "GRE Scores as Predictors of GPA for Psychology Graduate Students." *Educational and Psychological Measurement* 28 (2): 433–36.
- Nivet, Marc A. 2011. "Commentary: Diversity 3.0—A Necessary Systems Upgrade." *Academic Medicine* 86 (12): 1487–89.
- Orfield, Gary, ed. 2001. *Diversity Challenged: Legal Crisis and New Evidence*. Cambridge, MA: Harvard Publishing Group.
- Page, Scott. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ: Princeton University Press.
- Pallas, Aaron. 2001. "Preparing Education Doctoral Students for Epistemological Diversity." *Educational Researcher* 30 (5): 6–11.
- Pfeifer, Michael, and William Sedlacek. 1971. "The Validity of Academic Predictors for Black and White Students at a Predominantly White University." *Journal of Educational Measurement* 8 (4): 253–61.
- Posselt, Julie, Ozan Jaquette, Rob Bielby, and Michael Bastedo. 2012. "Access without Equity: Longitudinal Analyses of Institutional Stratification by Race and Ethnicity, 1972–2004." *American Educational Research Journal* 49 (6): 1074–1111.

Toward Inclusive Excellence in Graduate Education

- Roach, Ronald. 2005. "Affirmative Action Fallout." *Diverse: Issues in Higher Education* 22 (11): 24–27.
- Rojstaczer, Stuart, and Christopher Healy. 2012. "Where A Is Ordinary: The Evolution of American College and University Grading, 1940–2009." *Teachers College Record* 114 (7): 1–23.
- Smith, Daryl G. 2009. *Diversity's Promise for Higher Education: Making It Work*. Baltimore: Johns Hopkins University Press.
- Stake, Robert E. 2005. *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Stampnitzky, Lisa. 2006. "How Does 'Culture' Become 'Capital?' Cultural and Institutional Struggles over 'Character and Personality' at Harvard." *Sociological Perspectives* 49 (4): 461–81.
- Sternberg, Robert, and Wendy Williams. 1997. "Does the Graduate Record Examination Predict Meaningful Success in the Graduate Training of Psychologists? A Case Study." *American Psychologist* 52 (6): 630–41.
- Stevens, Mitchell. 2008. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, MA: Harvard University Press.
- Tierney, William, and Estela Bensimon. 1996. *Promotion and Tenure: Culture and Socialization in Academia*. Albany, NY: SUNY Press.
- Twombly, Susan. 1992. "The Process of Choosing a Dean." *Journal of Higher Education* 63 (6): 653–83.
- US Census Bureau. 2008. "2008 American Community Survey 1-Year Estimates," http://www.census.gov/acs/www/data_documentation/summary_file.
- Walker, George, Chris Golde, Laura Jones, Andrea Conklin Bueschel, and Pat Hutchings. 2007. *The Formation of Scholars: Rethinking Doctoral Education for the Twenty-First Century*. San Francisco: Jossey-Bass.
- Weber, Max. (1922) 1978. *Economy and Society*. Vol. 1. Berkeley: University of California Press.
- Wechsler, Harold S. 1977. *The Qualified Student: A History of Selective College Admission in America*. New York: Wiley.
- Whitla, Dean K., Gary Orfield, William Silen, Carole Teperow, Carolyn Howard, and Joan Reede. 2003. "Educational Benefits of Diversity in Medical School: A Survey of Students." *Academic Medicine* 78 (5): 460–66.
- Willingham, Warren. 1974. "Predicting Success in Graduate Education." *Science* 183 (4122): 273–78.
- Yosso, Tara. 2005. "Whose Culture Has Capital? A Critical Race Theory Discussion of Community Cultural Wealth." *Race, Ethnicity, and Education* 8 (1): 69–91.