Mya Poe, Norbert Elliot, John Aloysius Cogan Jr., and Tito G. Nurudeen Jr.

The Legal and the Local: Using Disparate Impact Analysis to Understand the Consequences of Writing Assessment

In this article, we investigate disparate impact analysis as a validation tool for understanding the local effects of writing assessment on diverse groups of students. Using a case study data set from a university that we call Brick City University, we explain how Brick City’s writing program undertook a self-study of its placement exam using the disparate impact process followed by the Office for Civil Rights of the US Department of Education. This three-step process includes analyzing placement rates through (1) a threshold statistical analysis, (2) a contextualized inquiry to determine whether the placement exam meets an important educational objective, and (3) a consideration of less discriminatory assessment alternatives. By employing such a process, Brick City re-conceptualized the role of placement testing and basic writing at the university in a way that was less discriminatory for Brick City’s diverse student population.

Demographically, the United States is changing. While people who identify as white, non-Hispanic on the US Census still outnumber other groups (63.4% to 36.6%), the birthrate for people who identify as African American, Native American, Hispanic, Asian, and mixed-race descent now totals more than 50% of all births in the United States (US, Census). This remarkable demographic change in the United States is having a profound impact on education. In some states, such as Texas and California, students of color now outnumber
white students in elementary and high schools (National Center). Similarly, at universities across the country, including Rutgers University in Newark, Stanford University, and many of the colleges in the CUNY system, students of color outnumber white students (Rutgers; Stanford; City University). And at institutions like California State University, Long Beach, Hispanic students are now the largest group on campus, comprising 30.8% of the undergraduate population (California State).

Such demographic changes mean that composition researchers must continue to understand how writing assessment practices impact diverse student populations (Haswell and Haswell; Hamp-Lyons; Inoue and Poe, *Race and Writing Assessment*). For example, we know that students of color are more likely to experience the negative effects of assessment because of rigid institutional requirements (Sternglass; Soliday). We also know that different writing assessment practices may yield quite different results that, in turn, yield different consequences (Kelly-Riley; Inoue). And when it comes to placement testing, we know that enrollment in noncredit, basic writing courses may either support or impede student writers. On one hand, basic writing may be an important, supportive environment for first-year students (Horner), as a number of advances have been made in the last two decades to help struggling writers in ways that situ-ate basic writing students positively within institutional structures (Glau). On the other hand, remedial identity remains defined, in large part, by a model of writing assessment mired in a narrow vision of writing (Condon), including highly constrained lexico-grammatical interpretations of use (Shapiro)—that is, decontextualized drills in grammar and usage. Moreover, students assigned to remedial courses may resist being required to take additional courses, may not enroll in those courses, may complete them at lower rates, and may graduate at much lower rates than their peers (Complete College; Scott-Clayton). Thus, validation tools that help us understand the local effects of writing assessment can be critically important: use of these tools may mean the difference between college success and failure for many students.

As articulated in the assessment literature, the definition of validity has shifted over time (Kane, “Validation”). Currently, the measurement community has forwarded a definition of validity based on interpretation and use of test scores (Kane, “Validating”). In their acknowledgment that use of assessment information is at the center of validity, measurement researchers have advanced
four sources of validity evidence in the call for validation studies: scoring (establishment of test taker’s performance through a scoring framework); generalization (evaluation of the different conditions that impact student performance), extrapolation (inference linking the test to a range of performances associated with the concept under investigation), and consequence (anticipation of intended outcomes, adverse impact, and systemic effects). Likewise, the composition community has begun to employ validation studies in a variety of contexts, including studies of how raters’ own language backgrounds and academic experiences influence decision making (Wiseman), design of integrated writing tasks for English language learners (Knoch and Sitajalabhorn), relationships between directed self-placement methods and curricular success (Gere et al.), and impact on ethnic groups of a portfolio system used to evaluate student readiness for upper-division writing (Kelly-Riley).

Although composition scholars have employed a variety of validation approaches, we have yet to investigate the usefulness of legal heuristics in validation research. Besides their obvious value in assessing legal responsibilities, legal heuristics remind us that decisions about assessment practices may be made through disciplinary frameworks outside of composition and measurement.

One particularly worthwhile legal heuristic for writing assessment research is the disparate impact approach, a legal analysis to determine unintentional discrimination. Although uncommon in the composition scholarship, disparate impact has received some attention in the education literature. For example, the foundational publication *Educational Measurement*, now in its fourth edition, includes an entire chapter dedicated to legal considerations in assessment (Brennan). Yet much of the legal scholarship in educational journals and measurement publications is problematic, focusing on federal law and failing to take state law into consideration (Sireci and Parker; Davis; Camilli; Pollock; Ryan; Kidder and Rosner), or containing incorrect statements of the law (Sireci and Parker; Phillips and Camara). As well, scholarship is often dated and therefore fails to consider important legal developments, such as relevant Supreme Court decisions (Sireci and Parker; Elul; Green) or more recent lower federal court decisions on disparate impact issues (Kidder and Rosner; Mank).

For composition scholars, however, the most disappointing aspect of this body of scholarship is its failure to recognize the disparate impact approach as
a validation tool rather than simply a means to understand legal responsibilities. This conceptual gap is disappointing because the legal approach offered through disparate impact analysis complements the pursuit of validity in writing assessment practice in two important ways. First, with its emphasis on the use of writing assessment scores, disparate impact analysis provides a robust entry point into complex economic, social, cognitive, and affective contexts that shape assessment scenes (Scott and Brannon). In other words, legal heuristics such as disparate impact analyses keep us focused on the relationship between scores and the local contexts in which decisions about assessments are made. Second, because of the tight connection in legal approaches between score use and context, evidence regarding scoring, generalization, and extrapolation is not to be understood as a discrete set of categories. Rather, each is understood in terms of its integrated impact on diverse student groups. Without contextualization of the evidence supporting the use of scores, the validation process is suspect.

In this article, we provide a model of how disparate impact analysis can function as a validation tool for understanding the effects of writing assessment practices on diverse groups of students. We use a data set drawn from a postsecondary institution located in Newark, New Jersey, that we call Brick City University. Specifically, our case study design, which is in the tradition of Breland and Ironson and their application of empirical models for fairness in law school admissions, employs data from Brick City to illustrate how the university’s writing program would undertake a self-study process of its assessment practices using the disparate impact process followed by Office for Civil Rights (OCR) of the US Department of Education, the federal agency charged with investigating disparate impact claims in educational settings. We do not propose the use of this method as a means to assess a writing program’s level of compliance with federal or state laws, a task best left to the university’s in-house legal department. Our aim is to advance a disparate impact process as a transparent method for self-study to accompany the use of an evidence-based approach to decision making that relies on quantitative information as well as contextualized reasoning. Although we have limited our discussion in this article to race, the disparate impact approach in the legal world extends beyond race and can be applied to other classifications, such as national origin, gender,
international students, and students with limited English proficiency. Thus, for the purposes of a validation study, the disparate impact approach that we describe here may be applied to any discrete, nonmajority group.

**Bias, Fairness, and Disparate Impact in Testing**

As a beginning, it is useful to define the terms *bias* and *fairness* and to explain their relationship to disparate impact. In composition studies, the term disparate impact is not often used; the term *bias*, however, is used frequently in the literature. In composition studies, elimination of bias is often used synonymously with *fairness* and has been associated with scoring consistency (White), ethics (Lynne), democratization (Broad et al.) and equity (Huot). Huot, for example, has argued that “it is time we visited fully the impact of assessments upon minorities, so that instead of adjusting test results, we could use tests that are fair to all” (9).

In large-scale educational measurement, where decontextualized testing is the norm, the *Standards for Educational and Psychological Testing* provide four principal ways in which the term *fairness* is used: lack of bias; equitable treatment in the testing process; equality in testing outcomes; and the opportunity to learn (American Educational Research Association; see also Baldwin). In addition, measurement researchers consider bias a technical term supported with statistical evidence. In defining test bias, for example, Cleary proposed the following: “A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent non-zero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score is consistently too high or too low for members of the subgroup” (115). Put simply, if a group’s actual performance is better than that group’s predicted performance, educational measurement researchers would say the test “underpredicts” the scores for that group; conversely, if a group’s performance is poorer than predicted, then the same researchers would say the test “overpredicts” scores for that group (Bridgeman, Pollack, and Burton). Additionally, for measurement researchers, mean differences in scores are insufficient ways to detect bias because specific abilities are so varied that it is impossible to believe that all populations are identical on all abilities (Reynolds), especially when different populations have different access to the very resources needed to develop abilities such as writing (Bridgeman).

In the legal context, fairness is a theoretical principle (Rawls), while determination of test bias usually focuses on evidence of discrimination—either intentional or unintentional—in specific cases. Discrimination in the legal
context is a technical concept and may be considered a constitutional or a statutory/regulatory concern at either the federal or the state level.

Determination of intentional discrimination, or *disparate treatment*, requires direct evidence of discriminatory intent; that is, evidence that there has been a purposeful effort to treat groups differently. In such cases, the intent of the test giver is the focus of the inquiry because the test giver’s discriminatory intent drives the test.

Another way of showing discrimination is by demonstrating unintentional discrimination, or *disparate impact*. Policies and practices that include no explicit racial or ethnic criteria and were never intended to discriminate can nevertheless have the same effect as overt discrimination (Bernal). We can, then, define disparate impact discrimination as the unintended racial differences in outcomes resulting from facially neutral policies or practices. Because discrimination flows from the test design, process, or use of test scores, rather than from the intent of the test giver, disparate impact analysis focuses on the *consequences* of specific testing practices.

For the purposes of understanding how legal scholars understand disparate impact and disparate treatment as technical concepts and to correct the partial and incorrect legal information that is currently found in the literature, we provide the following discussion of federal and state (New Jersey) constitutional and statutory/regulatory definitions of discrimination. This background is meant to help composition studies faculty understand the disciplinary context in which disparate impact legal analysis was developed and is used today, and this discussion, therefore, can help administrators explain the difference between the legal use of disparate impact and its use as a validation tool. In the legal context, disparate impact may only be used in specific kinds of cases (Table 1). As a validation tool, it may be used broadly for the purposes of self-study.

Federal antidiscrimination laws include, but are not limited to, the Equal Protection Clause of the Fourteenth Amendment to the US Constitution, Title VI of the Civil Rights Act of 1964, and Department of Education regulations implementing Title VI. State laws such as the New Jersey Law Against Discrimination (New Jersey Statutes) may provide legal protection in one state but will...
not be applicable in another state. For instance, both the US and New Jersey Constitutions guarantee “equal protection” under the law (State v. O’Hagen; US Const., Amend. XIV, sec. 1; Williams), which means that a state actor, such as Brick City University, must apply its laws and policies in an even-handed manner. However, such protections do not apply where there is no evidence of intentional discrimination (Greenberg v. Kimmelman; Washington v. Davis). That is, they cannot be used in cases of disparate impact. Furthermore, the equal protection doctrine only applies to discrimination by a state or by a state entity, such as a public university. It does not apply to a private university, even if the private university engages in intentional discrimination (Kaplin and Lee; Powe v. Miles; Shelley v. Kraemer).

Table 1. Which Antidiscrimination Laws Apply to Universities?

<table>
<thead>
<tr>
<th>Institution Type</th>
<th>Constitutional Claims</th>
<th>Statutory/Regulatory Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Federal</td>
<td>New Jersey</td>
</tr>
<tr>
<td>Equal Protection Clause of the 14th Amendment</td>
<td>Equal Protection under NJ state constitution</td>
<td>Title VI of the Civil Rights Act of 1964&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Institution Type</td>
<td>Federal</td>
<td>New Jersey</td>
</tr>
<tr>
<td>Applicable against public institutions</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Applicable against private institutions</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Type of Claim</td>
<td>Intentional discrimination</td>
<td>Yes</td>
</tr>
<tr>
<td>Disparate impact</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

a. Applies only to recipients of federal funds.
b. The New Jersey Law Against Discrimination does not apply to private religious educational institutions.
While constitutional avenues for disparate impact claims are limited, statutory or regulatory legal avenues such as Title VI of the Civil Rights Act of 1964, Department of Education regulations implementing Title VI, and New Jersey’s Law Against Discrimination provide some protection against disparate impact discrimination and apply more broadly than the equal protection doctrine. Since the analysis is similar for both state and federal laws, we focus here only on Title VI.

The applicability of Title VI of the Civil Rights Act of 1964 to disparate impact discrimination depends on whether the Title VI action is based on one part of Title VI (Section 601) or on Department of Education regulations issued under another part of Title VI (Section 602). Because this critical distinction has been very poorly explained in the assessment literature, we feel it is important to carefully distinguish applicability here.

Title VI was enacted to address racial segregation using Congress’s spending power (Watson). Section 601 of Title VI establishes this central claim: “No person in the United States shall, on the ground of race, color, or national origin, . . . be denied the benefits of, or be subjected to discrimination under any program . . . receiving Federal financial assistance” (42 US Code Sec. 2000d). Thus, Section 601 of Title VI imposes the same nondiscrimination obligation on all institutions that receive federal funds and, therefore, applies to all colleges and universities—public and private—that receive such funds.

Title VI does not, however, define what constitutes “discrimination” and does not specify whether the statute proscribes only intentional discrimination or whether it also reaches other forms of discrimination, such as those that produce disparate effects (Abernathy). The Supreme Court, however, has provided some direction on these issues. First, the Supreme Court has interpreted Section 601 as creating an implied private right of action (Cannon v. University of Chicago). This means that individuals can bring a discrimination lawsuit under Section 601 of Title VI against a university—either public or private—that receives federal funds. However, like the Equal Protection Clause of the US Constitution, Section 601 of Title VI has been interpreted to apply only to intentional discrimination (Guardians Association v. Civil Service Commission of New York City). No disparate impact claim can be brought under Section 601.

Section 602 has traditionally been the source of Title VI’s disparate impact enforcement standards. Section 602 authorizes federal agencies that disperse federal funds, such as the Department of Education, “to effectuate” the provisions of Title VI by issuing interpretive regulations (42 US Code Sec. 2000d-1). Like other federal agencies, the Department of Education has issued regulations
under Section 602 that implement Title VI. Those regulations specify that a recipient of federal funds may not use “criteria or methods of administration which have the effect of subjecting individuals to discrimination because of their race, color, or national origin” (34 Code of Federal Regulations Sec. 100.3(b)(2)). These regulations have been interpreted to prohibit disparate impact discrimination (Alexander v. Choate; Guardians Association v. Civil Service Commission of New York City).

For a time, the Supreme Court allowed individuals to file a disparate impact lawsuit under regulations that grew out of Section 602 (Lau v. Nichols). In 2001, however, the Supreme Court foreclosed the possibility of such private lawsuits in its Alexander v. Sandoval decision, when it held that Section 602 of Title VI does not create a private right of action. This effectively limited Title VI lawsuits to private or non-governmental cases involving intentional discrimination.

The Sandoval decision has been regarded as a major setback in addressing discrimination. However, the Supreme Court did not stop federal agencies, such as the Department of Education, from using their regulations as a basis to take action against disparate impact discrimination (Alexander v. Sandoval; Abernathy). Thus, the Department of Education’s Office for Civil Rights (OCR), the division charged with enforcing Title VI regulations, can still address disparate impact discrimination through its own investigations and administrative proceedings (West-Faulcon).

Finally, before continuing, we feel it is necessary to acknowledge that disparate impact analysis in the legal context is not without its critics. The disparate impact approach is contested—mainly because of theoretical debates about the nature of unintentional discrimination and the role of the courts in abating discrimination.

The disparate impact approach is contested—mainly because of theoretical debates about the nature of unintentional discrimination and the role of the courts in abating discrimination (Abernathy; Selmi). Some scholars argue that disparate impact is an inappropriate response to racial disparities (Braceras; Wax); some members of the Supreme Court have expressed skepticism regarding the theory of disparate impact (Ricci v. DeStephano). Despite these criticisms, however, the disparate impact approach has historically been an important tool for analysis in civil rights claims, and the Obama administration appears committed to the use of the disparate impact approach where applicable, having recently issued new regulations authorizing the use of the disparate impact theory to enforce the Fair Housing Act (US, Housing). As a result, the disparate impact approach to addressing racial disparities will likely remain viable under federal law, un-
less Congress makes this approach unlawful or the Supreme Court declares the disparate impact approach unconstitutional as a violation of federal Equal Protection standards. For the purposes of disparate impact self-study, we use the OCR investigation process as it is a clearly articulated process for self-study.

**Brick City University: A Portrait**

With terminology defined and legal precedent established, we turn now to the Brick City University case study context to explain how the disparate impact process may be applied in analyzing data from the university’s writing program.

Brick City University is a public four-year, doctorate-granting institution in Newark, New Jersey. In our case, the entering fall class totals 844 students. White students comprise about 39% of the enrolled Brick City student body, Hispanic students comprise 24%, Asian students comprise 23%, African American students comprise 13%, and Native American students comprise the remaining 1%. The mean SAT writing score of all groups of enrolled Brick City students is above average compared to state \((M = 499, SD = 116)\) and national \((M = 488, SD = 114)\) averages (College Board, State; College Board, Total; Mattern, Camara, and Kobrin).

The Brick City writing program is committed to local placement testing, with its WPA noting that the SAT writing test has “scant, limited validity” (Condon 104), in its measurement of a student’s ability to develop and support a point of view using the conventions of standard written English (College Board, Essay). Although not as progressive as directed self-placement or portfolio assessment, the Brick City writing program has attempted to make its placement exam meaningful, aligning the test and its curriculum with the WPA Outcomes Statement and using the placement exam as a tool for teacher training. The timed, impromptu exam, taking two hours, includes a writing task based on a reading passage in which students are required to write a persuasive essay and is scored by two readers, yielding a holistic score on a 6-point rubric designed to capture argument, organization, and knowledge of conventions. Following best practices in holistic scoring (White), readers are trained before the scoring, and scores differing by more than one point are read by a third reader. Adjudicated scores are combined to yield a final score ranging from 2 to 12. With the help of general education committee members, cut scores are
established. In the year of the case study, a combined score of 7 was established as the cut score for basic writing. Hence, a student who earned a score of 6 or below was placed into basic writing, a non-credit-bearing course. A score of 7 or above was used to place students into first-year writing. Students earning the highest scores could petition to be placed into honors first-year writing. Table 2 shows the placement scores for each of the admitted groups of students. Although there was no evidence of any intent to discriminate against certain groups of students, the Brick City Writing Placement Exam painted a weak portrait of student writing performance for African American and Hispanic students. As in a normally distributed population in which a histogram of the scores approximated the shape of a bell, nearly half of African American students (47%) and over one quarter of Hispanic students (28%) scored 6 or below and were placed in basic writing. Given that the placement rates reflected a normal distribution, the high rates of placement were disquieting to Brick City administrators. What should they make of these placement rates?

Of additional concern to the administration and the writing program administrator are the effects of remediation on student engagement and on student writing performance. Table 2 shows the placement scores for each of the admitted groups of students.

Table 2. Placement Exam Descriptive Statistics for Brick City University Admitted Students (n=844)

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean writing placement test scores</th>
<th>Basic writing (number and percent)</th>
<th>First-year writing (number and percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>6.86 ( SD = 1.48 )</td>
<td>50 (47%)</td>
<td>57 (53%)</td>
</tr>
<tr>
<td>Native American</td>
<td>7.33 ( SD = 1.5 )</td>
<td>2 (22%)</td>
<td>7 (78%)</td>
</tr>
<tr>
<td>Asian</td>
<td>7.87 ( SD = 1.41 )</td>
<td>29 (15%)</td>
<td>162 (85%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>7.42 ( SD = 1.35 )</td>
<td>57 (28%)</td>
<td>143 (72%)</td>
</tr>
<tr>
<td>White</td>
<td>7.99 ( SD = 1.22 )</td>
<td>35 (10%)</td>
<td>302 (90%)</td>
</tr>
</tbody>
</table>
graduation rates. While Asian and white students at Brick City University have graduation rates of about 59% and 54% respectively within six years, only about 40% of African American students graduate within six years, and only about 47% of Hispanic and Native American students graduate within six years. With such different graduation rates for certain groups of students, especially African American and Hispanic students, Brick City administrators worried that higher remediation rates were contributing to lower graduation rates. Was the placement test creating a barrier to graduation?

**Brick City’s Disparate Impact Self-Study**

The most practical way for Brick City to conduct a disparate impact self-study was to follow the “burden shifting” approach that is used in OCR investigations (US, Education, “Compliance”; US, Justice, “Title VI”). This approach involves a three-step inquiry:

- **Step 1:** Do the assessment policies or practices result in an adverse impact on students of a particular race as compared with students of other races?
- **Step 2:** Are the assessment policies or practices necessary to meet an important educational goal?
- **Step 3:** Even in situations where a college or university can demonstrate that the assessment policies or practices are necessary to meet an important educational goal, are there comparably effective alternative policies or practices available that would meet the school’s stated educational goal with less of a burden or adverse impact on the disproportionately affected racial group?

We now turn to each of these steps in greater detail.

**Step 1: Is there an adverse impact on certain students?**

The first requirement for making a disparate impact claim under the burden-shifting approach is to show an adverse impact on students of a particular race as compared with students of other races. In the context of its placement test, Brick City would begin its self-study through a statistical analysis of the pass
rates within the population of test takers, disaggregated by race. But what kind of statistical analysis should Brick City employ? While the educational measurement field uses standard statistical measures to determine the potential bias of a test, the legal world relies on no one measure for a disparate impact analysis (West-Faulcon). However, two approaches that are often used are the “four-fifths rule” or a test of statistical significance (Peresie). Under the “four-fifths rule,” disparate impact is found when the effects of a testing policy or practice on a particular race have effects of less than 80%, or four-fifths, of the rate of effects on the reference group (West-Faulcon). If Brick City were to use the four-fifths method, self-study leaders would determine if the use of scores from its placement test—disaggregated by race—resulted in certain groups of students being placed in basic writing at a rate lower than 80% of the placement rate of white students, given that white students had the highest rate of placement into first-year writing.

According to its data, Brick City places 90% of white students in first-year writing but places only 85% of Asian students, 78% of Native American, 72% of Hispanic students, and 53% of African American students in first-year writing. In order to meet the four-fifths rule, Asian, Native American, Hispanic, and African American students would have to be placed in first-year writing at a rate of at least 72%—that is, 80% of the rate of white students (.8 * .9 = .72). While the placement of Asian, Native American, and Hispanic students meets the four-fifths rule because they place into first-year writing at 85%, 78%, and 72%, respectively, African American students place into first-year writing at a rate of only 53%, well below the 72% minimum to meet the four-fifths rule. Hence, the writing placement test at Brick City violates the four-fifths rule for African American students and would be a prima facie case of disparate impact. In other words, this finding points to a presumed case of disparate impact under Step 1 of the three-step burden-shifting approach. This finding also suggests that Brick City must either change its test design, process, or use of test scores from the placement exam or justify its exam under Steps 2 and 3.

In addition to a four-fifths analysis, Brick City could use a test of statistical significance, such as chi-square analysis, to search for evidence of disparate impact. One such test, the Pearson chi-squared test, is used to determine whether distributions of categorical variables—variables that can only have one of two possible values (in our case, placement into either basic writing or first-year writing)—differ from one another in a statistically significant way. The chi-squared value and the resulting p-value explain the likelihood that the disparities between placement rates based on race are due to chance.
Replication of the approach used by West-Faulcon reveals that the \( p \)-values for the basic skills placement rates of white students compared to Hispanic students \( (\chi^2 (1, n = 537) = 29.01) \) and white students compared to African American students \( (\chi^2 (1, n = 444) = 63.9) \) are both less than 0.0001. That is, the differences are extraordinarily statistically significant, meaning that there is a strong likelihood that the differences in placement rates for Hispanic and African American students were due to the effect of the placement test. Tables showing both the four-fifths analysis and the chi-squared analysis of Brick City University’s Writing Placement results are available on the CCC website (www.ncte.org/cccc/ccc/issues/v65-4).

Which test should Brick City use, the four-fifths rule or a test of statistical significance for the purposes of self-study? It depends. The four-fifths approach is mathematically simpler and therefore easier to apply as an initial “rule of thumb” approach (US, Equal Employment). The chi-squared test, on the other hand, is a more robust statistical measure, identifying placement differences that are statistically significant, and statistical significance in this case implies that the differences in placement are not due to chance alone but instead may be due to the placement test itself. ⁶

Ultimately, the presence of a significant disparity, shown by either the four-fifths rule or a test of statistical significance, is only the beginning of the analysis. A placement test that causes a significant disparity in placement rates for different groups of students is not necessarily problematic under the disparate impact analysis if the test can be justified by an educational necessity.

**Step 2: Is there a justifiable educational need?**

To identify the educational necessity of remediation determined through placement testing, the writing program should undertake a multiphase inquiry. To identify the educational necessity of remediation determined through placement testing, the writing program should undertake a multiphase inquiry. To identify the educational necessity of remediation determined through placement testing, the writing program should undertake a multiphase inquiry. To identify the educational necessity of remediation determined through placement testing, the writing program should undertake a multiphase inquiry. To identify the educational necessity of remediation determined through placement testing, the writing program should undertake a multiphase inquiry. To identify the educational necessity of remediation determined through placement testing, the writing program should undertake a multiphase inquiry.

To identify the educational necessity of remediation determined through placement testing, the writing program should undertake a multiphase inquiry. The inquiry should include the following procedures: determining whether the elements of writing that the test measures are important components of student writing ability; ensuring that the test results in fact capture those components; documenting that the basic writing course provides help to students for the identified writing deficiencies; and demonstrating that the test is predictive of or significantly correlated with students’ performance in college writing. Each phase of the study would need to be supported with empirical evidence.
While the Brick City writing program has identified elements of writing important to student success, namely argument with emphasis on an integrated approach to technical reading and writing, and has specified that the placement test capture those components, predictive studies would take at least one semester to demonstrate empirically if the basic writing course remediated the identified writing deficiencies and as long as six years to determine if the placement test was predictive of or significantly correlated with students’ performance in terms of graduation rates (Student Right-to-Know).

One way to establish the ability of students to perform the written aspects of college work at an acceptable entry level is for the Brick City writing program to see how students who are placed into basic writing end up faring in the traditional, credit-bearing course offered the very next semester. Upon examination, for example, the Brick City writing program could survey instructors in basic writing and first-year writing to see if they believe students are accurately placed by the exam. As well, analysis of student course performance might also demonstrate that students who are placed into basic writing and receive a grade of B or better in that course also receive grades of B or better when they matriculate to first-year writing. However, if further analysis reveals that many students who receive a B or better in basic writing receive grades of D or F in first-year writing (and those students were sufficiently homogenous in educational backgrounds that they would be expected to perform consistently across both courses), then an additional analysis would be needed—this time of the basic writing curriculum. Examination of instructor comments on student texts might reveal that despite writing instructor training, grades were in fact being awarded on the basis of knowledge of conventions and that the other elements of the curriculum—argument and organization—are undervalued. So, while the placement exam was identifying students who needed assistance with developing college-level arguments, the basic writing curriculum was not supporting the development of such abilities. Brick City administrators would then have to make a decision on how to handle writing placement in future admitted classes. Evidence of adverse impact had been determined, and the argument for justifiable educational need had been
qualified by examination of the basic writing curriculum and student performance in subsequent courses. At this point, a one-semester study design—that is, a curricular experiment—would prove extremely useful.

**Step 3: Are alternatives available that meet the school’s stated educational goal with less of a burden?**

A unique benefit of a legal heuristic is that it encourages us to imagine a wide range of alternative assessment processes. Ultimately, following its self-study, the Brick City writing program should logically choose a method of assessment that would result in equally good outcomes for all its students but without the racialized score distribution of its previous timed, impromptu exam and subsequent student performance. In the Brick City case, the best solution was twofold: (1) changing the timing of the placement exam to better align it with classroom instruction and (2) mainstreaming all students into first-year writing. In this final phase of the disparate impact analysis, demonstration that the placement test is predictive of or significantly correlated with students’ performance in college writing over a six-year period could now be undertaken in an environment in which the threat of disparate impact had been justifiably lessened because the educational need for the placement test had been realized in the curriculum. In this process, Brick City administrators could design a comprehensive first-year curriculum in which the cognitive, intrapersonal, and interpersonal elements of writing identified in the *Framework for Success in Postsecondary Writing* were used (O’Neill et al.). The original placement assessment was maintained but given within the first two weeks of the semester as a formative measure: students with low scores would be advised for additional support. Because Brick City administrators now adopted the proposition that any admitted student was qualified to begin credit-bearing coursework, a less discriminatory alternative was made available.

**Consequence: A Way Forward**

Contemporary demographic changes in the United States usher in important challenges for writing assessment researchers who are committed to issues of fairness in the assessment of writing. As our analysis suggests, disparate impact analysis not only provides a tool for analyzing placement rates for different groups, but it also encourages us to articulate our educational reasons for such placements ... it pushes us to consider possible alternative measures as part of the validation process.
analysis provides promising directions for self-study as a validation tool. Disparate impact analysis not only provides a tool for analyzing placement rates for different groups, but it also encourages us to articulate our educational reasons for such placements. As a validation tool, it connects testing and curriculum. Most importantly, it pushes us to consider possible alternative measures as part of the validation process. In doing so, it expands how we think of the local in assessment practice. As we have demonstrated, while locally developed measures are vital, they are not a proxy for validity.

The first step for writing researchers to begin a self-assessment study focusing on placement under the disparate impact approach is to calculate the placement rates at their institutions, using the methods we have provided here, according to race, gender, and other subpopulations (certainly, at a minimum, those covered by applicable federal and state disparate impact laws). The second step is to explicate the educational need of a particular assessment practice. In this step, if a disparate impact is found, then a more contextualized inquiry must be undertaken in order to understand whether the test meets an important educational objective. The third step, one we believe must be undertaken concurrently with the data disaggregation and analysis, is to imagine a less discriminatory assessment alternative. Such reconceptualization is complex, and we recognize that the preservation of the rich culture of basic writing must be balanced with the gains of mainstreaming (McNenny and Fitzgerald; Otte and Smoke). We are not advocating for the elimination of basic writing in all cases. In the Brick City case, that alternative was available because of the kinds of support that the university had available to commit to first-year writing.

As the Brick City case reveals, the three-step disparate impact process yields distinct advantages for Brick City students, teachers, and administrators.

As the Brick City case reveals, the three-step disparate impact process yielded distinct advantages for Brick City students, teachers, and administrators. First, the process yielded an assessment of writing that was fully situated within a unified curriculum for all students. Second, it provided teachers and administrators confidence in the placement process as the problems with the previous placement test would be eliminated. Finally, because time to graduation will be a factor in the federal government’s recent Postsecondary Institution Rating System planned as part of the 2014 reauthorization of the Higher Education Act, the writing program will no longer be part of the process of remediation associated with diminished federal allocation of student aid based on institutional graduation rates (US, Education, “Request”).
In the end, making good decisions about our writing assessment practices for all students means attending to the various ways that we understand the impact of assessment on our students. Disparate impact analysis is an important self-study tool for the composition community—a tool that speaks to our historical and current struggles to provide meaningful writing instruction to all students. Multidisciplinary teams such as the one formed by the authors of this paper reveal the potential for validation studies in which composition, legal, and quantitative methods inform each other to suggest ways that we might make writing assessment fairer for all students.

Acknowledgments
We would like to thank Brent Bridgeman, Chris Gallagher, and Neal Lerner for their comments on earlier versions of this article. We would also like to thank Keith Rhodes, Peggy O’Neill, and Irwin Weiser for their insightful reviews.

Notes
1. Descriptive and inferential statistical models for the Brick City case were drawn from a population of admitted students to demonstrate patterns of ethnic diversity and associated student performance reported in the literature of writing assessment (e.g., Elliot et al.); current and proposed policies associated with the Family Educational Rights and Privacy Act (FERPA) are therefore not applicable.

2. A regulation, which is different from a statute, is a detailed rule promulgated by an administrative agency, such as the US Department of Education, under authority granted to the agency by a statute. Regulations outline how statutes will be interpreted and applied by an administrative agency. Both statutes and regulations have the force and effect of law.

3. Legal cases focusing on disparate impact can give rise to criticisms of so-called reverse discrimination, or intentional discrimination against whites. For example, in the classic case of employment discrimination, a group of underrepresented firefighters who are denied promotion because of written examination results seek to challenge the examination using disparate impact analysis. Because there are a limited number of slots for promotion, the de facto result is that one group of firefighters (those who were denied promotion) are pitted against other firefighters (those who received promotions and stand to lose those promotions if the examination is successfully challenged) (Plaut; Norton and Sommers; Norton). These problems do not arise when disparate impact analysis is used as a writing assessment self-study tool because there should be no stated quotas or established limits defined by policy on the number of students who are placed in basic writing. Thus, there is no “zero-sum” scenario where students are pitted against each other.
4. Because the number of Native American students is typically small in the admitted class, institutional researchers suggested that any longitudinal studies would be of individual students—not of students as representative of a larger group.

5. In the case of an OCR investigation, this approach would unfold as follows. If an investigation reveals that the Brick City’s placement exam results in significant disparities between groups based on race (Step 1), the burden shifts to Brick City to articulate why the assessment policies or practices are necessary to meet an important educational goal (Step 2). If Brick City fails to identify the important educational goal served by the discriminatory assessment policies or practices, the institution will be in violation of Title VI. If Brick City can identify the important educational goal served by the discriminatory assessment policies or practices, it will nevertheless be in violation of Title VI if there are less discriminatory, comparably effective alternative policies or practices available that meet the stated educational goal (Step 3).

6. In addition to the four-fifths rule and the Pearson chi-square test, other appropriate statistical measures may also be useful in validation analyses that require more nuanced approaches. In such cases, the determination about which test should be used is best developed in consultation with an expert in statistics.

Works Cited


Braceras, Jennifer C. "Killing the Messen-


Bridgeman, Brent. Message to the authors. 29 Mar. 2013. Email.


Frank v. Ivy Club. 120 NJ 73. Supreme Court of NJ. 1990. Print.


———. “Validation.” Brennan 17–64.


———. Dept. of Education. “Request for Information to Gather Technical Expertise Pertaining to Data Elements, Metrics, Data Collection, Weighting, Scoring, and Presentation of a Post-secondary Institution Rating System.”


Watson, Sidney D. “Reinvigorating Title VI: Defending Health Care Discrimina-


---

**Mya Poe**

Mya Poe is assistant professor of English at Northeastern University. Her scholarship includes *Learning to Communicate in Science and Engineering: Case Studies from MIT* (with Neal Lerner and Jennifer Craig), *Race and Writing Assessment* (with Asao B. Inoue), and a special issue of *Research in the Teaching of English* featuring international perspectives on assessing diverse writers.

---

**Norbert Elliot**

Norbert Elliot is professor of English at New Jersey Institute of Technology. With Edward M. White and Irvin Peckham, he is coauthor, most recently, of *Very Like a Whale: The Assessment of Writing Programs*.

---

**John Aloysius Cogan Jr.**

John Aloysius Cogan Jr. is associate professor of law and Roger S. Baldwin Scholar at the University of Connecticut School of Law.

---

**Tito G. Nurudeen Jr.**

Tito G. Nurudeen Jr. was a member of the Albert Dorman Honors College at the New Jersey Institute of Technology, earning a BS in science, technology, and society. While attending NJIT, he interned for the Office of the Attorney General, Division of Consumer Affairs. Research related to the case study in this paper was supported by the provost’s Undergraduate Summer Research Award. He will attend Rutgers–Newark School of Law in 2014.