Journal of Writing Assessment

Del Principe, Annie (2023). Time as a "Built-In Headwind" : The Disparate Impact of Portfolio Cross-Assessment on Black TYC Students. *Journal of Writing Assessment,* 16(1). **DOI**: 10.5070/W4jwa.1452

Time as a "Built-In Headwind"
The Disparate Impact of Portfolio Cross-Assessment on Black TYC Students

Annie Del Principe, Kingsborough Community College, CUNY, US, Ann.DelPrincipe@kbcc.cuny.edu

Abstract: This study of a departmental portfolio cross-assessment practice sheds light on factors that appear to influence assessment outcomes for Black students and helps to tease out some of the reasons why this assessment ecosystem has a disparate impact on these students. The findings, drawn from student outcomes data and student survey data, suggest that it isn't only, or even primarily, Black students' linguistic variety that leads to higher failure rates. The writing qualities most commonly flagged on Black students' failing portfolios are likely related to the very different material conditions in which they write their papers. These conditions challenge the framing of "time" and "labor" as neutral, non-racially-inflected resources to which all students have equal access and that are not often conceptualized as part of the construct of writing ability. As TYCs across the country reform their placement mechanisms for greater access and equity and place more and more students of color into their credit-bearing FYC courses, we have an ethical obligation to watch for disparate impacts created by our preexisting assessment ecosystems.

Keywords: disparate impact, Black or African American students, portfolio assessment, intersectionality, time, TYC writing assessment

In 2009, a group of faculty in the English Department in my institution, Kingsborough Community College (KCC), CUNY, developed our own faculty-driven, homegrown version of collaborative portfolio assessment (CPA) across many sections of our required core FYC 1 course. These faculty were active in writing studies fields and knew that portfolio assessment is considered the gold standard in writing assessment, as it relies on multiple samples of student writing, and often several different drafts or iterations of those writings, and allows assessors to consider a body of work from a student rather than basing judgment on a single piece of student writing (Huot, 2002; White, 2005). The faculty also knew that communal cross-assessment of student writing portfolios is considered by many writing assessment scholars to be the most authentic and effective way to create a flexible and negotiated consistency in assessment across multiple sections of a required course (Broad, 2000, 2003; Colombini & McBride, 2012), and the faculty drew on Bob Broad's (2003) practice of "dynamic criteria mapping" to develop a local approach to communal assessment that reflected their experience and values. For many years, participation in CPA was voluntary on the part of the faculty member teaching the course. If the faculty member opted in, then all of their students would be assessed at the end of the term via CPA. If they didn't opt in, then they would assess their own students' performances in whatever way they chose. Over time, CPA grew in popularity in our department, and when—like many other colleges in the nation—the CUNY system began to develop co-requisite versions of FYC 1 and eliminated timed testing as a final assessment hurdle in those co-requisite courses, the department opted to implement CPA across all co-requisite sections of FYC 1. This expanded tremendously the number of sections participating in this assessment practice, and by 2019-20, the year this study data was collected, this meant that fully one-third of all sections of FYC 1 were assessed via CPA.

Initial disaggregated outcomes data after the department made the switch from timed-final assessment to CPA was extremely encouraging and made faculty feel quite positively about the change in assessment practice. Student pass rates overall increased after the switch to CPA, and disaggregated outcomes data showed that, in particular, Hispanic, Asian, and male students (all races/ethnicities) experienced higher outcomes.[1] However, faculty were disappointed and puzzled by the fact that Black students' pass rates didn't also rise after the switch to CPA; the rates didn't fall, but they also did not rise like the rates for other minoritized groups. The significant differences in pass rates in CPA for students from different racial groups inspired the current study, which strives to understand how and why these disparate outcomes are being produced. More specifically, by looking closely at teacher feedback and student survey input, this study sheds light on the factors that appear to influence assessment outcomes for Black students and helps to tease out some of the reasons why an assessment ecosystem like CPA has a disparate impact on these students. The findings suggest that it isn't only, or even primarily, negative assessments of Black students' linguistic variety that lead to higher failure rates in CPA. Instead, it seems as if Black students at KCC experience higher failure rates via CPA due to concerns about other qualities of their writing performance rather than "mechanical correctness" or "clarity." Further, the writing qualities most commonly flagged on Black students' failing portfolios are likely related to the very different material conditions in which they write their papers for FYC 1. These conditions challenge the framing of "time" and "labor" as neutral, non-racially inflected resources to which all students have equal access and which are not often conceptualized as part of the construct of writing ability.

1 The racial category labels used in this article—Black, Hispanic, and Asian—are being used because they are the categories used by the CUNYfirst student information system, from which some of the data in the study was drawn.

As TYCs across the country reform their placement mechanisms for greater access and equity and place more and more students of color into their credit-bearing FYC courses, we have an ethical obligation to watch for disparate impacts created by our preexisting assessment ecosystems in those courses.

Black Students & Writing Assessment: A Turbulent Relationship

For quite a long time, writing teachers and writing assessment scholars have known that BIPOC students tend to score lower on various types of writing assessments (Breland et al., 2004; Camara & Schmidt, 1999). One common hypothesis for this racial variation is that, because BIPOC students are frequently also from lower socioeconomic backgrounds, they have had weaker educational preparation and backgrounds as compared to White students. This deficit framing acknowledges that there are racialized variations in assessment but locates the cause, or the "problem," safely outside the assessment itself; from this perspective, the cause of the racial inequity is the racial inequity of the broader society, and the assessment simply reflects the students' inequitable educational preparation.

Students' use of non-standard varieties of English, and assessors' biases against those varieties, is also often cited as the cause of these lower scores and lower pass rates. Assessors, operating under the "myth of linguistic homogeneity" (Matsuda, 2006), have traditionally treated criteria such as "clarity" and "mechanical correctness" as neutral qualities of written English by which student writing can be evaluated, perhaps resulting in the lower assessments assigned to BIPOC students' writing given the non-standard varieties of English these writers often employ. Further, the issue of language variation doesn't end with syntax and grammar; it also involves cultural practices involving texts that influence the ways different writers and readers interpret and interact with texts (Murphy, 2007).

For Black students in particular, scholars have studied and discussed ways of understanding their lower achievement on standardized writing assessments as a consequence of their use of African American Vernacular English (AAVE) and of the assessors' bias against this particular variation of English. Geneva Smitherman's (1992, 1993) prodigious scholarship has documented the grammatical, syntactical, and discourse-level qualities of spoken and written AAVE and has analyzed the impact the use of these features in writing has on assessment decisions. Through Smitherman's repeated close analyses of writing samples from African American students on the NAEP tests, it has become clear that students' use of AAVE grammar, syntax, and discourse have changed in frequency over time (with AAVE grammar and syntax declining since 1969) and that these three language features do not necessarily correlate with each other—meaning, writing that includes more of one feature does not necessarily include more of the others (Smitherman, 1993). Further, in her studies, Smitherman (1993) found that writing samples that had a higher degree of AAVE discourse style markers had fewer instances of AAVE grammar and syntax markers, suggesting that "'talking Black' does not have to encompass features of BEV grammar" (p. 17). In general, the more markers of AAVE discourse in a writing sample, the higher the rating, but, importantly, this was only true in the "imaginative/narrative" writing prompt and was not true in the "persuasive" writing prompt. In response to this last finding, Smitherman calls, repeatedly, for writing teachers and assessors at all levels of education to make changes in their instruction, including to "design strategies for incorporating the Black imaginative, story-telling style into

student production of other essay modalities" (p. 21) and to encourage writing from a more "field-dependent" style, which she'd found was a more comfortable writing style for Black students.

Decades ago, one of the factors that motivated many writing teachers and scholars to move away from timed single writing assessments (e.g., the NAEP and SAT) and to develop the rich practice of performance-based portfolio assessment was their recognition of the inequities and the invalidity of the writing construct implicit in the timed approach. Still, despite portfolios' very different approach to assessment and different framing of the writing construct, we have continued to see racial disparities in student outcomes. Diane Kelly-Riley's (2011, 2012) work provides the closest analysis of a situation like the current study—a cross-assessment practice for writing portfolios in a large institution—and reports on the racial disparities she found in the student outcomes. When a Black student commented to Kelly-Riley that she heard Black students failed the university-wide writing portfolio at a higher rate than other students, Kelly-Riley (2012) was inspired to investigate whether "the shared evaluation processes used by the university-wide Writing Portfolio assessment—and by other contextually defined writing assessment practices—be inadvertently complicit in perpetuating a system of discrimination? In other words, could teacher/evaluators unwittingly be disadvantaging students of color in a large-scale writing assessment program because of unstated biases related to race?" (p. 3). Kelly-Riley found that Black students did perform significantly lower on the portfolio review than all other students. In attempting to tease out whether this was the result of unstated bias related to race, Kelly-Riley analyzed the assessment criteria used by assessors and found that while student assessments did vary along racial categories, these did not seem to be a result of the scorers/raters using racially/demographically charged criteria to score student writing. She drew this conclusion because issues of "coherence" were more influential than "mechanics" in raters' assessments of African American students' portfolios, making the assumption that the only racially inflected criteria were on the sentence-level surface of the writing and that raters' judgements of "coherence" might not be indirectly influenced by racially inflected discourse and rhetorical styles (p. 11).

The focus on surface-level AAVE features to explain the pattern of disparate outcomes for Black students doesn't consider the contributions of two other important factors: 1) AAVE, and all language variations, influences language behavior at levels other much deeper than grammar and syntax; and 2) racial inequities are frequently caused by the less-obvious structural racism deeply embedded in our assessment practices in addition to the more obvious causes. Scholars including Arnetha Ball, Carmen Kynard, and Geneva Smitherman—to name but a few—have described the preferred written discourse styles of Black students whose work they've studied and have called attention to the ways that these styles differ from the "field independent" (Smitherman, 1993) "essayist literacy" (Farr, 1993) typically anticipated in writing assessments. Smitherman observed Black students received higher scores on "imaginative/narrative" writing prompts than they did on "persuasive" writing and that they preferred to use a "field dependent" style that relies more heavily on cultural influences and conversational tone. Ball (1996) identified that variations in AAVE aren't only on the level of grammar and syntax but also include "styles of rhetoric" such as "implicit linking of topics in the discourse, shifts in focus, and topic relationships which must be inferred by the message receiver" (p. 226). Because portfolio raters might not recognize the rhetorical and discourse style variations in students' work as linked to their racial identity, they run the risk of "disrupt[ing] the rich potential of African Diasporic discourses and relegat[ing] students' critical logos, ethos and pathos as un/non-academic" (Kynard, 2008, p. 5).

In this current study and analysis, I'd like to expand our understanding of what racialized assessment practices might look like by reframing them in light of antiracist conceptions of structural racism. Since the Black Lives Matter protests of the summer of 2020, the antiracist movement has become widely known and discussed in our society and in our discipline. Of course, decades before 2020, critical race theorists tried to draw attention to the myriad ways in which our legal systems manage to appear race-neutral while simultaneously allowing the continuation of the racially inequitable structures that lie below the surface in our society. More recently, Eduardo Bonilla-Silva (2017) offered the concept of "color-blind racism," a racial ideology that rationalizes endemic economic, political, and social disparities among racial groups as nonracial, as a way to understand how we are able to perpetuate racism without feeling ourselves to be racists. Ibram X. Kendi (2019) contributed his how-to manual to the conversation, asserting that "antiracist" refers to "one who is expressing the idea that racial groups are equals and none needs developing, and is supporting policy that reduces racial inequity" (p. 24). Although there are important differences among these different waves of scholarship, the common thread is the point: institutional practices that perpetuate racial inequity in our society are frequently missing the distinctive, obvious racist packaging that was openly accepted in earlier historical periods. Many practices in our society that appear to be fair are grounded in assumptions rooted in deep, long-standing racial inequities.

Bringing this conversation back to writing assessment and our quest to understand why Black students at KCC continue to receive lower scores and ratings on their work, the conceptual shift from surface-level racism to structural racism requires that we consider how various aspects of the assessment ecosystems we've created perpetuate racially inequitable results. We know that assessing students for "correctness" has a disproportionately negative effect on Black students and other students who do not speak or write Standard American English, and we've come to suspect that other commonly-used criteria— clarity, coherence, and structure, to name a few—might also not be racially neutral in the way we imagined they were. The contract grading or labor-based grading movement has tried to design a way around these racist realities by stopping the use of qualities of writing and language to assess writing performance and ability. Instead, many teachers and assessors who practice contract or labor-based grading create agreements with their classes regarding the quantity of work or labor—articulated in number of words, number of projects, and/or amount of time—that students must complete to earn a certain grade in the course. Although grading contracts are not the same as writing portfolios (the focus of the current study), they have a significant similarity that makes them relevant to this discussion: in both approaches to writing assessment, students must produce a certain, often substantial, amount of writing in order to participate successfully in the assessment practice. Asao Inoue (2019), a well-known scholar and proponent of labor-based grading, names one of the central goals of labor-based writing assessment as "cultivating with our students an ecology, a place where every student, no matter where they come from or how they speak or write, can have access to the entire range of final course grades possible" (p. 3). By using labor as the fairest common denominator by which all students, regardless of their raciolinguistic backgrounds, will be judged, Inoue (2019) aims to "dismantle White language supremacy in schools and society" (p. 4).

Despite labor-based grading's overt emancipatory goal of racial equity, it hasn't proven able to fully escape having a negative impact on Black students' grades. In order to better understand how grading contracts worked for students from different racial backgrounds, Inoue (2012) studied the experiences and outcomes for students in a composition course that used grading

contracts and analyzed their responses to an end-of-term survey, their final portfolio ratings, and their grade distributions. Inoue (2012) found that, while contract grading did not seem to harm Black students—because their scores were similar to the mean scores for all students— "their grade distributions achieve fewer 'As' and 'Bs' than any other group . . . suggesting that they had trouble meeting the quantity required by the contract and getting the work done" (p. 89). Further, more Black students than any other group did not meet the minimum requirements of their contract in terms of the quantity of the work produced (Inoue, 2012, p. 90). Inoue's (2019) later analysis of the performance of his own students using his grading contracts showed the same pattern with most students of color in the middle and bottom grade groups (p. 252). Ellen Carillo (2021) points out that while labor-based grading may resolve some inequities related to raciolinguistic diversity, it perpetuates other inequities that it doesn't acknowledge. Carillo argues that labor-based grading may "unintentionally privilege some students over others" (p. 8) because it "inaccurately assumes that labor is a neutral measure—or at least that it is less inequitable a measure than quality" (p. 11).[2]

Labor-based grading's struggle to produce racially equitable assessment results might be somewhat puzzling, given the fact it effectively eliminates racialized variations in language as a possible trigger for variations in assessment. However, this situation is only puzzling when we assume that the most salient racialized elements of students' writing are linguistic and skill-based variations. Instead, when we broaden our understanding of the intersectional realities of race and socioeconomic status, and when we broaden our understanding of the writing construct to include material conditions and resources that are unequally distributed due to long-standing structural racism, the persistence of Black students' apparent underperformance via our assessments is much less puzzling. It is by stepping back from a narrower focus on the "trees" of language variation and including more factors in our conceptualizations of the multiple, intersectional writing ecosystems within which we write that we can gain a clearer view of the entire forest of variables that contribute meaningfully to what we think of as writing ability. The current study helps to illuminate two important factors, very likely the results of the intersection of socioeconomic status and raciolinguistic identities, that seem to influence Black students' performance on their portfolios.

Method for this Study

Within the fields of writing studies and writing assessment, there has been growth in attention to disparate impact analyses over the last several years. Poe et al. (2014) introduced the use of "disparate impact," a concept borrowed from law, as a validation tool in writing assessment. They describe a process by which a sample institution, "Brick City University," could undertake a self-study of its own writing placement exam to investigate whether it resulted in indirect discrimination against Black students by employing the process used by the Office for Civil Rights of the U.S. Department of Education when investigating disparate impact claims. This current study uses and broadens the "step 1" of the Office of Civil Rights process—"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?" (Poe et al., 2014, p. 599)—in its examination of the

2 Since 2012, Inoue's theory and practice of assessment has evolved from a "hybrid" approach that quantifies work produced by students but saves the highest course grades for those whose writing is judged to be high quality to a labor-based approach that presents three nuanced dimensions of labor that do not include judgements of writing quality.

disparate impact on Black students of a final assessment, rather than a placement, practice. After establishing that disparate impact is occurring (see "wave 1" below), this study tries to understand which elements of the assessment ecosystem seem to be contributing to that disparate impact (see "wave 2" and "wave 3" below). My hope is that, by making visible how the CPA practice produces disparate impact for Black students, this study will help render this "site of oppression . . . actionable" (Kryger & Zimmerman, 2020) and will inspire and require our institution to take meaningful action to create a fairer assessment system for our students.

Three waves of data collection and analysis were necessary in order to flesh out enough details of Black students' experiences with CPA to be able to begin to discern factors connected to their lower pass rates. Because this study is not only trying to indicate a disparate impact situation but is also trying to illuminate why that disparate impact might be occurring, it was important to gather and bring together multiple types of information.

Wave 1: Disaggregated Pass Rates in FYC 1

In order to construct a clear picture of overall student outcomes in FYC 1 in Fall 2019, I analyzed racially disaggregated data provided to me by Kingsborough's Office of Institutional Effectiveness on course outcomes (course completion, pass rates, and grades) for all students in sections of English 12 that implemented collaborative portfolio assessment (CPA). The racial categories used in Kingsborough's internal student data system—Black, Hispanic, Asian, White, and Native American—are pre-determined and controlled by CUNY. When students register in the university system, they select one or more racial or ethnic (Hispanic/non-Hispanic) categories along with other demographic information. This disaggregated course outcomes data allowed me to set my study data within its broader context and to understand whether there was, in fact, disparate impact on certain student groups.

In determining whether there was disparate impact on student groups, I employ the simple "four-fifths rule" (West-Faulcon, 2009) to disaggregated course pass rates for FYC 1. The four-fifths rule establishes that disparate impact is found if the results of a policy have a resulting effect of 80% (four-fifths) or less on a particular group as compared to the reference group. In the current study, this means that I compare the racially disaggregated course pass rates of student groups and see whether any or all minoritized student groups—Black, Hispanic, and Asian—have pass rates that are 80% or lower than the pass rates of White students. This analysis confirms a basis for a more detailed look at the factors that might lead to the production of different pass rates for Black students in particular.

Wave 2: Disaggregated Assessment Criteria Applied to Black Students

In order to discern whether there were patterns correlated with student race in the writing traits cited as weak by faculty assessors, I gathered copies of 92 feedback forms written by faculty portfolio assessors for portfolios that failed via the CPA process. During the period of this study, final student portfolios were required to include two pieces of writing that students had worked on in the course—including early drafts of these pieces with instructor and/or peer feedback—plus a final self-assessment essay in which students were asked to provide portfolio readers with a "tour" of their writing samples from the course and point out ways in which those writings provided evidence that they had demonstrated growth and proficiency in the final assessment criteria. I asked faculty teaching and assessing these courses to share all feedback forms with me, and I

estimate that I received about 75% of all forms generated via the CPA process. The feedback forms listed the student's CUNY student ID number but not their actual name. I then coded the feedback forms for racial identity by using students' records on the CUNYfirst student data system. Of the 92 students whose portfolio feedback forms I had collected, 76 had listed a racial identity in the student information system. Of the 16 students who did not list a racial identity, 8 listed their ethnicity as Hispanic. Coding the feedback forms by race allowed me to determine whether there were meaningful differences in the ways Black students' failing portfolios were assessed in comparison with the portfolios submitted by non-Black students.

Wave 3: Disaggregated Student Survey Responses

In addition to these differentiated assessment results, I sought to capture the attitudes and experiences of students from different racial backgrounds regarding the entire process of CPA. To this end, I distributed two slightly different surveys to all students enrolled in the 25 sections of FYC 1 participating in CPA assessment in Fall 2019. The first survey was distributed on week 10 (in a 12-week term) of the semester, and the second survey was distributed after the end of the semester and once all final grades had been posted. I felt it necessary to survey the students before the semester ended in order to capture their attitudes towards CPA and their experiences preparing their writing for CPA before those attitudes were influenced by their result (pass or fail) via the assessment. The second survey (after the semester had ended and grades were in) aimed to gather students' attitudes and experiences toward CPA from the perspective of having finished the course and knowing the results of their assessment.

The surveys asked students to identify the racial group(s) with which they identify and then asked them to respond to a series of multiple choice and open-ended free-response questions regarding how they felt about CPA, how prepared they felt for CPA, and whether they thought CPA was a fair assessment process. In addition to these questions targeting attitudes towards CPA assessment, the surveys asked students to share additional demographic information: gender identity, English as a first language status, work status, and caregiving status. The post-course survey asked them a similar set of demographic questions followed by a series of multiple choice and open-ended free-response questions regarding whether they passed CPA, factors they think affected their CPA outcome, and whether they think CPA is a fair assessment process.

I distributed these Google forms surveys via email and text message to all students enrolled in all 25 sections of FYC 1 implementing CPA assessment in the Fall 2019 term, a group that totaled 625 students at that time. 106 students responded to the first survey, and 69 students responded to the second survey. I exported the results of the surveys to Excel and to Hyperresearch, where I was able to analyze the responses for differentiated patterns based on racial group identification.

Results

Wave 1: Disparate Impact in FYC 1

In Fall 2019, 850 total students were enrolled in sections of FYC 1 that implemented CPA. Of those students, 669 "completed" the course, which means that they did not officially withdraw by the withdrawal date and that they received a grade for the course. Note that completing the course does not necessarily mean that a student submitted a final portfolio. I do not have access to an accurate count of how many of the 669 actually submitted a final portfolio. Of the 669 students

Table 1

Relative frequency distribution of student outcomes by race in fall 2019 FYC 1 courses using CPA assessment

Ethnicity	students in sample		% pass	% failure	% incomplete portfolio
	n	%			
NA/PI	1	0.1	0.0	0.0	100
Black	285	33.5	43.2	28.8	28.1
Hispanic	169	19.8	58.6	21.3	20.1
Asian	141	16.5	60.3	18.4	21.3
White	254	29.8	71.3	14.6	14.2

completing the course, 57.4% passed and 21.3% failed. Given the focus of the current study, it is illuminating to review a graphic of these outcomes disaggregated by race (see table 1).

The disaggregated outcomes display quite well the racially disparate impact that the current study is trying to understand with greater nuance and detail. We can see quite easily the relatively lower pass rates for Black students as compared to all other racial and ethnic groups.

When we apply the four-fifths rule (West-Faulcon, 2009) to these outcomes, we see that Black students' pass rates are 60.5% of White students' pass rates. According to the four-fifths rule, disparate impact can be considered to have occurred if Black students' pass rates are 80% or less of White students' pass rates. Therefore, these data suggest disparate impact has occurred in this assessment practice. When establishing disparate impact in any assessment situation, it is important to always remember that disparate *impact* is not the same thing as disparate *treatment* (Moreland, 2018, p. 188). This analysis does not claim that Black students were treated differently than other students during the assessment process. Instead, this analysis is emphasizing the uneven consequences of this seemingly-neutral assessment practice—final writing portfolios—for Black students. This uneven profile in pass rates is typical of outcomes data we have seen over the past several semesters in this course, and this persistent inequity in outcomes is what inspired the current study. In order to understand how the disparate impact is functioning to disadvantage Black students and to understand some of the factors that might contribute to it, we'll look closely at the data from portfolio feedback forms and from student survey responses.

Wave 2: Criteria Flagged in Failing Portfolios

32 of the 76 racially identified students in the study identified themselves as Black in the CUNYfirst student data system when they enrolled in the college. Overall, criteria cited as weak in these Black students' failing portfolios (see table 2) followed general trends for non-Black students' failing portfolios (see table 3). "Analysis" was the most frequently cited criterion, followed in order of frequency by "focus," "organization & structure," and then by "development & growth" and "requirements of the assignment" (see table 2). 40.6% of Black students' failing portfolios cited "mechanical correctness" as one of the criteria leading to failure, placing that criterion as the 6th

Table 2
Relative frequency distribution of criteria cited as weak on Black students' failing portfolios

Criterion	Rate of Citation (%)
Focus	56.25
Analysis & critical thinking	71.87
Reading	28.12
Organization & structure	56.25
Research	18.75
Mechanical correctness	40.63
Meets requirements of assignment	53.13
Development and growth	53.13
Incomplete portfolio	12.5

Note. *N*=32 portfolios.

Table 3
Relative frequency distribution of criteria cited as weak on non-Black students' failing portfolios

Criterion	Rate of Citation (%)
Focus	65
Analysis & critical thinking	75
Reading	26.7
Organization & structure	61.7
Research	26.7
Mechanical correctness	43.4
Meets requirements of assignment	48.3
Development and growth	61.7
Incomplete portfolio	3.3

Note. *N*= 60.

most commonly cited criterion for these students. Fully 12.5% of failing Black student portfolios failed because they were incomplete. This is notable, as it is much higher than the rate of incomplete portfolios from any other student racial grouping.

As compared to other student racial categories of color—Hispanic and Asian—Black students had a similar profile of criteria cited for failure. For these three groups of students, "analysis & critical thinking" and "focus" were the first two most commonly cited criteria for

triggering failure on the portfolio. As is true for Black students, Hispanic students' next two most commonly cited criteria were "organization & structure" and "development & growth." While "mechanical correctness" was cited on 40.6% of Black students' failing portfolios, it was cited on 23.8% of Hispanic students' portfolios, making this criteria less commonly cited for Hispanic students. Interestingly, "mechanical correctness" was cited on 75% of Asian students' failing portfolios, making it as frequently cited as "focus" as a cause for failure. This makes Asian students the student group most likely to have had their portfolios failed due to concerns about their linguistic correctness or purity, which is an analysis worthy of a different article.

As compared to White students who failed the portfolio assessment, Black students had a very different profile of criteria for failure. In the order of frequency, the following criteria were flagged on failing portfolios of White students' work: "development & growth," "organization & structure," "analysis & critical thinking," and "focus." Notably, 47% of White students' portfolios had "mechanical correctness" listed as a criterion for failure, which is higher than the 40% of Black students' portfolios having that same criterion named. These findings echo those of Kelly-Riley (2011, 2012) who found that "mechanics" was not frequently cited as the cause of failing portfolios for Black students in her study. Of course, when reviewing these data, it is important to remember how many more Black students fail FYC 1 and fail via the CPA process than do White students. 28% of Black students failed FYC 1 in Fall 2019 as compared to the 14.6% of White students who failed. Still, this marked difference in the criteria cited as weak by assessors and, therefore, leading assessors to recommend failure in the course is clearly racially linked and deserves closer analysis.

Wave 3: "It's not fair because the requirements is so much and we only have a little bit of time to get them done" (Student 34)

106 students responded to the survey I circulated in week 10 of the semester, and patterns in the responses from Black students help to further contextualize this group's portfolio results. The 106 respondents represent a 16.9% response rate based on the total student population assessed via CPA in Fall 2019 and who were still enrolled in the course at week 10 of the semester. The racial breakdown of respondents loosely mirrors the enrollment of the course (see table 4).

The key takeaways from the student surveys fell into two broad categories, the first of which is that Black students reported much less confidence about whether they would complete and pass

Table 4

Showing the racial breakdown of survey respondents and students enrolled in the course

Race/Ethnicity	Survey Respondents[a]		Course Enrollment	
	n	%	*n*	%
Black	27	25.4	285	33.5
Hispanic	23	21.6	169	19.8
Asian	22	20.7	141	16.5
White	41	38.6	254	29.8

[a]Seven student respondents are double-represented because they are multi-racial.

Table 5
Rate of Confidence of Course Completion by Race/Ethnicity

Race/Ethnicity	Confidence of course completion (%)
Black	48.1
Hispanic	82.6
Asian	77.2
White	78

Table 6
Rate of Confidence of Passing Course by Race/Ethnicity

Race/Ethnicity	Confidence of passing[a]
Black	2.6
Hispanic	3.3
Asian	3.75
White	3.73

[a] On a Likert Scale from 1 (*very uncertain*) to 5 (*very confident*)

their FYC course than did other student groups. Table 5 shows the disparity between Black students' rate of confidence that they will complete the course compared with other student groups. Course "completion" in this case means remaining enrolled in the course until the end of the semester and continuing to submit all required work. In an FYC course, this means completing a variety of writing projects. Further, given that these sections of FYC all engaged in CPA, completing the course means finishing at least three major writing projects including various drafts, or iterations, of those writings. Along with being less confident that they would complete the course, Black students were also less confident that they would pass the portfolio at the end of the course (see table 6). The data presented in table 6 was drawn from a Likert scale question asking students to rate their confidence they would pass the portfolio on a scale from 1 (very uncertain) to 5 (very confident).

The second key takeaway from the student surveys is that Black students reported having much busier schedules than did students in all other groups. Black students in this study were far more likely to work (59.20%) than most other groups and were much more likely to work full-time (33.30%) than all other groups (see figure 1 and table 7). In addition to larger work responsibilities outside of school, Black students in this study were also more frequently caregivers to either children or dependent family members as compared with other student groups. Fully 37% of Black

Figure 1

Relative Frequency Distribution of Student Work and Caregiving Rates (%) by Race/Ethnicity

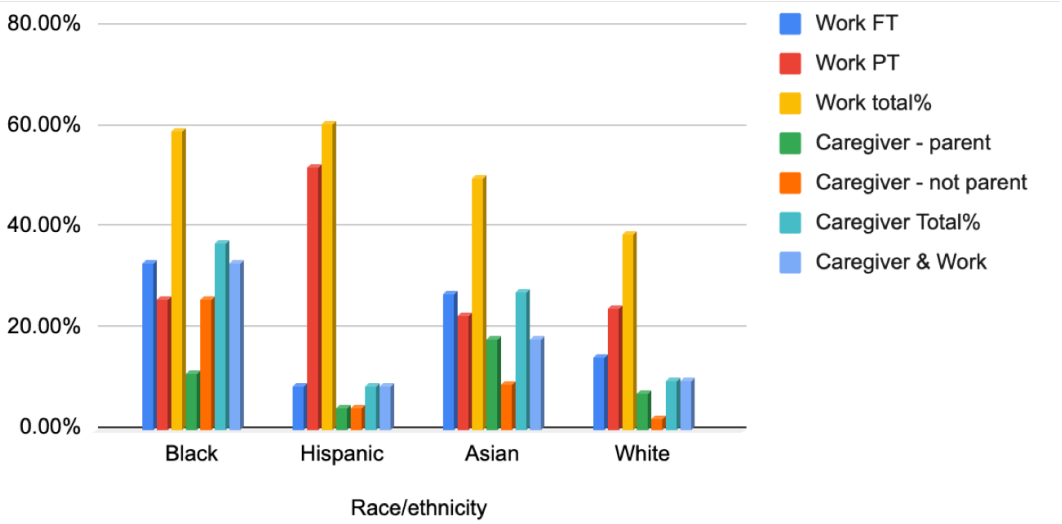

Table 7

Comparative Rates (%) of Work and Caregiving by Race/Ethnicity

Race/ Ethnicity	Work (full time)	Work (part time)	Work (any)	Caregiver (parent)	Care-giver (not parent)	Care-giver (any)	Work and Care-giver
Black	33.3	25.9	59.2	11.11	25.9	37.0	33.33
Hispanic	8.6	52.17	60.8	4.3	4.3	8.6	8.7
Asian	27.2	22.7	50.0	18.0	9.1	27.3	18.0
White	14.6	24.3	39.0	7.3	2.4	9.75	9.75

students reported being a parent or acting as a caregiver to a dependent family member, whereas students from other groups reported much lower caregiving rates: Hispanic 8.6%, Asian 27%, White 9.75%. When we look at how frequently Black students in this study both work outside of school and have a caregiving role, we find that a third (33.33%) of students play both of these roles.

Given the substantial amount of time required to complete writing assignments, it's quite possible that Black students frequently find themselves without enough time to complete their writing and frequently miss deadlines due to the much higher rates of time-consuming responsibilities—work and caregiving—they report having outside of school. We might expect to see the effect of this time crunch in lower pass rates and lower course completion rates for these students; and, in fact, we do see those lower rates in the broad course outcomes data. Being short on time may also connect to Black students' lower reported rate of ease in producing the amount of

Table 8
Ease of Producing Writing by Race/Ethnicity

Race/Ethnicity	Ease of Producing Writing[a]
Black	2.29
Hispanic	2.87
Asian	3
White	2.85

[a]On a Likert Scale from 1 (*very difficult*) to 5 (*very easy*).

writing required for their portfolios. Table 8 shows student responses to a Likert scale question asking them to rate their sense of ease in producing enough writing from 1 (very difficult) to 5 (very easy).

Overall, patterns in responses from Black students on the midterm survey put in relief the different levels of confidence these students have regarding their ability to complete and pass FYC 1 and the different levels of responsibility outside the classroom Black students in this study experience. This highlights the reality that time itself is a racially inflected, rather than racially neutral, resource and that inequities in the availability of time might be connected to Black students' lower outcomes in this course via CPA. As one Black student survey respondent wrote of the CPA process, "it's not fair because the requirements is so much and we only have a little bit of time to get them done" (student 34).

Discussion: Nuisance Variables, Intersectionality, and Structural Racism in Writing Assessment

> Fairness demands equal educational opportunities for all students including ethnic minorities, students with disabilities, low-income students, and English language learners. However, these subgroups of students have historically lagged behind their mainstream peers on test scores due, in part, to factors that may not be closely related to their academic achievement but do influence their performance outcomes. In order to begin reducing the performance gaps between subgroups and mainstream students, improvements must be made in both instruction and assessment by identifying and controlling for these factors (often referred to as nuisance variables) that affect a subgroup's academic achievement. (Abedi & Gandara, 2006, p. 43)

This study of the negative disparate impact of a portfolio assessment practice (CPA) on Black students suggests that it's time to reconsider the presumption of time as a non-racialized resource equally distributed to all students. Our field's shift away from timed writing assessments and toward portfolios, labor-based contracts, and similar approaches has helped improve the validity of the assessments, the learning opportunities embedded in the assessments, and, very likely, the fairness of the assessments. At the same time, we've tended to ignore the fact that assessment practices that demand a large quantity of writing and position the time required to produce that

writing as race-neutral, ignoring the intersectional reality of race and socioeconomic status in our society. For the Black students in this study, time appears to be one of the "nuisance variables" (Abedi & Gandara, 2006, p.43) or "latent variables" (Poe & Cogan, 2016) that lurks outside of the more obvious factors we imagine might create racially disparate results in portfolio assessment. Due to higher rates of caregiving and outside employment, Black students in this study simply have less time available to produce the large amount of writing required by portfolio assessment with fully one-third of these students reporting substantial work and caregiving duties. This may help explain why Black students had a much higher rate of incomplete portfolios and why their portfolios failed due to criteria that are likely linked to inadequate time: "development & growth," "organization & structure," and "requirements of the assignment."

Of course, time is far from the only other factor that affects the way Black students—or many other students—are rated in portfolio assessment systems. Various elements of students' positionalities come together and influence how any assessment system functions for them. Gender, age, trauma and mental health histories, and neurodiversity (Carillo, 2021), to name a few, will and always do affect the fairness of any form of writing assessment. And any one position cannot be assumed to always create the same influence for all students. For example, in the current study, there was no noticeable difference in ratings or survey responses connected to the gender of Black students, but there could easily be an important gender difference observed in a different study of Black TYC students or if we looked closely at students in this same study from a different racial background.

If we take seriously Poe, Nastal, and Elliot's (2019) challenge to consider all *admitted* students to be *qualified* students, then it is incumbent upon us to create not only placement mechanisms that are equitable, fair, and ethical but also assessment mechanisms and opportunity structures for these same students that do not ignore the profound diversity—in numerous domains—that the "least advantaged" TYC students bring to the table (Elliot, 2016). TYC placement reforms have resulted in a much higher number of students placing directly into credit-bearing FYC courses and, more significantly here, a much higher percentage of Black and otherwise minoritized students placing into those courses than occurred during the previous decades when these students would have been placed into developmental sequences. While much scholarship on how these students fare in FYC courses is positive and shows higher pass, completion, and retention rates than ever before, this is likely not true in all institutions. The current study is just one example of how inequities embedded in the construct of writing implicit in an assessment practice are surfaced by including a more and more diverse student population into the practice. By disaggregating student outcomes, by looking closely at who does and does not benefit in an assessment ecosystem, and by taking the time to gather empirical local data to inform our interpretations, we can begin to unpack the aspects of our assessments that are contributing to further inequity for our students. Here, the seemingly fair, normative resource of "time to write and revise papers" turned out to be strongly differentiated by race and made it more difficult for Black students to succeed.

The ultimate goal of studying situations of disparate impact, like the one in this study, isn't only to make writing assessments and outcomes more equitable but, more importantly, to make them more *fair*. In his presentation of a theory of ethics for writing assessment, Norbert Elliot (2016) centers fairness as the ultimate goal of writing assessment and defines it as "the identification of opportunity structures created through maximum construct representation. Constraint of the writing construct is to be tolerated only to the extent to which benefits are realized for the least

advantaged." In the current study, the writing construct had been broadened along some axes—multiple iterations of multiple different samples of student writing, including metacognitive self-assessment—and constrained along others—requiring a large amount of time outside of class available to produce numerous writing performances—and the least-advantaged group, Black students, did not realize a benefit from this version of the construct. To reference Chief Justice Berger's simile for the friction testing and diploma requirements created for Black job applicants in the *Griggs v. Duke Power Company* (1971) case, the requirement of such extensive writing outside of class acts as a "built-in headwind" for Black students as they attempt to meet the requirements of CPA and pass FYC 1 (qtd. in Poe & Cogan, 2016). In attempting to create a portfolio assessment ecosystem that was broad, flexible, and valid—in the sense that it best reflected "real" college writing—we, faculty at my institution, had unwittingly and unintentionally created a system that was racially biased against our Black students and very likely against all of our students from lower socioeconomic status, too.

While the purpose of this article is not to propose a clear assessment solution for our situation, I will suggest that our department employ the next two steps the Office of Civil Rights of the Department of Education uses to help guide its progress in cases of disparate impact. The first step, cited above, asks whether a policy or practice results in "adverse impact" on students of a particular race. As this article has shown, CPA does have a disparate impact on Black students in that a higher percentage of students fail credit-bearing FYC 1, thereby losing financial aid, momentum towards their degree, and confidence that they can succeed in higher education. Step 2—"Are the assessment policies or practices necessary to meet an important educational goal?" (cited in Poe et al., 2014, p. 599) —asks educational communities to reflect deeply on the underlying purposes and goals of the practice under scrutiny. Certainly, some type of assessment is productive and necessary across the 80+ sections of FYC 1 offered at my TYC every semester, but we must ask ourselves what aspects of cross-marked portfolio assessment seemed (when we designed them) and *still* seem vital to the fair and compassionate assessment of student learning in the course. When we initially designed CPA, we were aware of the truism that "assessment drives instruction" and wanted to create a process that both provided the best opportunity to students to showcase their strengths and their learning and that provided impetus to teachers to ask students to produce a fair amount of regular writing in their courses. In our experience, without that accountability, some teachers would slide into an approach to teaching the course that involved precious little actual writing on the students' part. But it seems that, in trying to make sure all students in the course actually wrote a certain quantity, we created a racist assessment ecosystem that made it harder for Black students to succeed.

If the answer to step 2 is a "yes" —the practice is necessary to meet an educational goal—the Office of Civil Rights' third and last step invites educational communities to imagine equally effective assessment alternatives to the practice. In undertaking this step, my department would need to rethink what sort of assessment practice would allow us to ensure that students have learned in the course, provide opportunities for students to name and showcase what they've learned, and also inspire strong instruction across the entire program. Certainly, we would aim to maintain the aspects of portfolio assessment—particularly cross-assessed portfolio assessment—that make it an effective and strong practice for assessing student learning while reducing the amount of available time outside of class our current CPA practice presumes of each student.

Conclusion

The purpose of this article is to shed light on some factors that appear to contribute to a pattern in outcomes for Black students that has been observed by many scholars. In particular, assessment ecosystems grounded in the production of a substantial quantity of writing—such as portfolio assessment or some approaches to labor-based assessment—might put Black students at a disadvantage due to the extensive time outside of class necessary to perform successfully on this type of assessment. Because race and socioeconomic status are strongly intersectional in the United States (Byrd et al., 2019), it should not be surprising that what feels like a factor related to socioeconomic status (free time) is also closely related to race. I join Carillo (2021) and Kryger and Zimmerman (2020) in drawing attention to the intersectional realities of students' lives and to ways in which assessment practices that focus on one axis of privilege "do not fundamentally intervene on other intersectional dynamics of power and privilege. Instead, they sidestep the deeply problematic and subjective quality-based assessment practices and exchange them for a less understood but still marginalizing focus on labor, performance, ability, and time" (Kryger & Zimmerman, 2020).

Further, I do not assume that only Black students are put at a disadvantage by these types of assessment systems. Any lower/working-class student, often a non-traditional TYC student, who is trying to balance caregiving and work responsibilities while attending school, would be put in the same position as a large percentage of the Black students in this study. As we try to develop writing assessments that do not presume that students have extensive time outside of class, we might find that assessment ecosystems that focus on what students have learned, rather than on the amount they have worked or have produced, might better avoid this problem. Of course, learning outcomes are quite tricky and would have to be developed in ways that allow all students to exhibit what they know and have learned about writing and about texts without positioning one particular racialized style of writing, or of work, as the norm. I, for one, believe this is possible.

Finally, a brief word about the lure of deficit thinking. In my 20+ years of experience as a TYC writing teacher and WPA, I have observed—in my own thinking and in the thinking of the numerous teachers I've worked with—that, when faced with a situation that appears to show disparate impact, the first move we make in explaining student failure is to say something like "they (the students) weren't ready for the course" or "they weren't adequately prepared for the course," and we point to the easily visible social inequities in our society to explain the patterns in student underperformance. Scholars do this, too, when they struggle to find other ways to explain differential outcomes for different groups of students, essentially washing their hands clean from any responsibility for reproducing social inequality. While it stands to reason that we cannot expect to always see precisely the same pass rates for all groups of students all the time, I believe we must change our orientation from accepting that the racial and social inequalities of the larger society will always be replicated in our assessment outcomes to the more radical, inclusive orientation that our assessments are likely part of that inequitable society and, like other structures, must be changed to advance fairness and justice. Even though "test scores may reflect social inequality . . . the use of test scores works to create that social inequality" (Poe & Cogan, 2016). We must look quite hard at *what* we are teaching and *what* we are assessing and be on the lookout for places where unquestioned assumptions lead to inequitable outcomes, as they did for Black students in this study. To do anything less is to wash our "hands of the conflict between the powerful and the

powerless . . . not to be neutral" (Freire, 1985, p. 122) and to perpetuate the structural violence of writing assessment (Lederman & Warwick, 2018).

References

Abedi, J., & Gandara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice*, 25(4), 36–46. https://doi.org/10.1111/j.1745-3992.2006.00077.x

Ball, A. F. (1996). Expository writing patterns of African American students. *The English Journal,* 85(1), 27–36. https://doi.org/10.2307/821117

Ball, A. F. (1999). Evaluating the writing of culturally and linguistically diverse students: The case of the African American Vernacular English speaker. In C. R. Cooper & L. Odell (Eds.), *Evaluating Writing: The Role of Teachers' Knowledge about Text, Learning, and Culture* (pp. 225–248). NCTE.

Bonilla-Silva, E. (2017). *Racism without racists: Color-blind racism and the persistence of racial inequality in America* (5th ed.). Rowman & Littlefield Publishers.

Breland, H., Kubota, M., Nickerson, K., Trapani, C., & Walker, M. (2004). *New SAT writing prompt study: Analyses of group impact and reliability* (Report No. 2004-1). College Entrance Examination Board. https://files.eric.ed.gov/fulltext/ED562781.pdf

Broad, B. (2000). Pulling your hair out: Crises of standardization in communal writing assessment. *Research in the Teaching of English*, 35(2), 213–260.

Broad, B. (2003). *What we really value: Beyond rubrics in teaching and assessing writing*. Utah State University Press.

Byrd, W. C., Brunn-Bevel, R. J., & Ovink, S. M. (2019). *Intersectionality and higher education: Identity and inequality on college campuses*. Rutgers University Press.

Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification*. Report No. 99-5. College Entrance Examination Board. https://eric.ed.gov/?id=ED562656

Carillo, E. (2021) *The hidden inequities in labor-based contract grading*. Utah State University Press.

Colombini, C. B., & McBride, M. (2012). "Storming and norming": Exploring the value of group development models in addressing conflict in communal writing assessment. *Assessing Writing*, 17(4), 191–207.

Elliot, N. (2016). A theory of ethics for writing assessment. *Journal of Writing Assessment*, 9(1). https://escholarship.org/uc/item/36t565mm

Farr, M. (1993). Essayist literacy and other verbal performances. *Written Communication*, 10(1), 4–38.

Freire, P. (1985). *The politics of education: Culture, power, and liberation*. Bergin & Garvey.

Huot, B. (2002). Toward a new discourse of assessment for the college writing classroom. *College English*, 65(2), 163–180. https://doi.org.kbcc.ezproxy.cuny.edu/10.2307/3250761

Inoue, A. B. (2012). Grading contracts: Assessing their effectiveness on different racial formations. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 79–94). P. Lang.

Inoue, A. B. (2019). *Labor-based grading contracts: Building equity and inclusion in the compassionate writing classroom*. The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PER-B.2019.0216.0

Kelly-Riley, D. (2011). Validity inquiry of race and shared evaluation practices in a large-scale, university-wide writing portfolio assessment. *Journal of Writing Assessment*, 4(1). https://escholarship.org/uc/item/7m18h956

Kelly-Riley, D. (2012). Getting off the boat and onto the bank: Exploring the validity of shared evaluation methods for students of color in college writing assessments. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 29–43). P. Lang.

Kendi, I. X. (2019). *How to be an antiracist* (1st ed.). One World.

Kryger, K., & Zimmerman, G. X. (2020). Neurodivergence and Intersectionality in Labor-Based Grading Contracts. *Journal of Writing Assessment*, 13(2). Retrieved from https://escholarship.org/uc/item/0934x4rm

Kynard, C. (2008). Writing while Black: The colour line, Black discourses and assessment in the institutionalization of writing instruction. *English Teaching: Practice and Critique*, 7(2), 4–34.

Lederman, J., & Warwick, N. (2018). The violence of assessment: Writing assessment, social (in)justice, and the role of validation. In M. Poe, A. B. Inoue, & N. Elliot (Eds.), *Writing assessment, social justice, and the advancement of opportunity*. The WAC Clearinghouse; University Press of Colorado. DOI: 10.37514/PER-B.2018.0155.2.07

Matsuda, P. K. (2006). The myth of linguistic homogeneity in U.S. college composition. *College English*, 68(6), 637–651.

Moreland, C. (2018). Chasing transparency: Using disparate impact analysis to assess the (in) accessibility of dual enrollment composition. In M. Poe, A. B. Inoue, & N. Elliot (Eds.), *Writing assessment, social justice, and the advancement of opportunity*. The WAC Clearinghouse; University Press of Colorado. DOI: 10.37514/PER-B.2018.0155.2.07

Murphy, S. (2007). Culture and consequences: The canaries in the coal mine. *Research in the Teaching of English*, 42(2), 228–244.

Poe, M., & Cogan, J. A. (2016). Civil Rights and Writing Assessment: Using the Disparate Impact Approach as a Fairness Methodology to Evaluate Social Impact. *Journal of Writing Assessment,* 9(1). Retrieved from https://escholarship.org/uc/item/08f1c307

Poe, M., Elliot, N., Cogan, Jr., J. A., & Nurudeen, Jr., T. G. (2014). The legal and the local: Using disparate impact analysis to understand the consequences of writing assessment. *College Composition and Communication*, 65(4), 588–611.

Poe, M., Nastal, J., & Elliot, N. (2019). Reflection. An admitted student is a qualified student: A roadmap for writing placement in the two-year college. *Journal of Writing Assessment,* 12(1). https://escholarship.org/uc/item/31c793kf

Smitherman, G. (1992). Black English, diverging or converging?: The view from the National Assessment of Educational Progress. *Language and Education*, 6(1), 47–61.

Smitherman, G. (1993, November 17-22). "*The blacker the berry, the sweeter the juice*": *African American student writers and the National Assessment of Educational Progress* [Conference

session]. The Annual Meeting of the National Council of Teachers of English, Pittsburgh, PA, United States. https://files.eric.ed.gov/fulltext/ED366944.pdf

West-Faulcon, K. (2009). The river runs dry: When Title VI trumps state anti-affirmative action laws. *University of Pennsylvania Law Review*, 157, 1075-1160.

White, E. M. (2005). The scoring of writing portfolios: Phase 2. *College Composition and Communication*, 56(4), 581–600.