

Chapter 11

How Administrative Data Collection and Analysis Can Better Reflect Racial and Ethnic Identities

SAMANTHA VIANO 

George Mason University

DOMINIQUE J. BAKER 

Southern Methodist University

Measuring race and ethnicity for administrative data sets and then analyzing these data to understand racial/ethnic disparities present many logistical and theoretical challenges. In this chapter, we conduct a synthetic review of studies on how to effectively measure race/ethnicity for administrative data purposes and then utilize these measures in analyses. Recommendations based on this synthesis include combining the measure of Hispanic ethnicity with the broader racial/ethnic measure and allowing individuals to select more than one race/ethnicity. Data collection should rely on self-reports but could be supplemented using birth certificates or equivalent sources. Collecting data over time, especially for young people, will help identify multiracial and American Indian populations. For those with more complex racial/ethnic identities, including measures of country of origin, language, and recency of immigration can be helpful in addition to asking individuals which racial/ethnic identity they most identify with. Administrative data collection could also begin to incorporate phenotype measures to facilitate the calculation of disparities within race/ethnicity by skin tone. Those analyzing racial/ethnic disparities should understand how these measures are created and attempt to develop fieldwide terminology to describe racial/ethnic identities.

Advances in the use of comprehensive administrative data sets have allowed researchers to answer significant questions focused on educational policy and practice (Connelly et al., 2016; Figlio et al., 2017). Mirroring these advancements is a growing literature that problematizes conventional racial and ethnic (R/E) categories by developing theoretical models that include additional layers of individual

Review of Research in Education

March 2020, Vol. 44, pp. 301–331

DOI: 10.3102/0091732X20903321

Chapter reuse guidelines: sagepub.com/journals-permissions

© 2020 AERA. <http://rre.aera.net>

identity (Ladson-Billings, 2012; Monroe, 2013; Pang et al., 2011), a movement spurred by the acknowledgment that R/E has been undertheorized in education research (King, 2016). While, arguably, those working on both of these advances in the literature have similarly equity-minded goals, combining these two approaches remains a challenge (Dixon-Román, 2017). Administrative data sets—information gathered about an entire population of individuals often collected by the government (Figlio et al., 2017)—tend to include a limited range of categories for R/E, and quantitative researchers often utilize even fewer of these categories in analysis (e.g., combining smaller R/E groups into an “other” category) to create parsimonious models (Denton & Deane, 2010; Ladson-Billings, 2012). At the same time, some critical race theorists, building on a theory that has been discussed for over a hundred years (DuBois, 1899; Jones, 1998),¹ suggest that quantitative analysis is unsuitable for studying inequality and outcomes based on R/E due to the history of the development of statistics in conjunction with racist movements like eugenics (Covarrubias & Vélez, 2013; Gillborn et al., 2018; Zuberi & Bonilla-Silva, 2008).

Prior research on the use of R/E in the social sciences has often focused on the actual collection and categorization of R/E data (e.g., Denton & Deane, 2010) and how participants are categorized into R/E groups, with less attention paid to how researchers then use those categorizations in their analyses. In this synthesis, we compile the knowledge and insights from this literature to further the field’s understanding of how to measure R/E in administrative data and then analyze these data to understand trends and disparities by R/E. For the purposes of this study, we use the definition of administrative data used by Figlio et al. (2017). Administrative data sets in education are collected by schools across the K–16 pipeline (a) that include a census of all students (and possibly employees) in that school or institution, (b) that are collected for administrative purposes, and (c) with the school, institution, or their management organizations “owning” the data (though researchers can apply for access). Since administrative data are a census of all students, they provide the opportunity for additional analysis focused on all students beyond the capacity of analysis of survey data. Collecting and analyzing administrative data about R/E are important for education researchers and administrators to be able to address the unique needs of different student groups, particularly in light of the persistent disparities in the experiences and outcomes of certain students.

The goal of this chapter is to create a resource for both researchers with access to administrative data and practitioners managing administrative data systems. While quantitative researchers tend to utilize R/E measures from administrative data sets without recognizing the flaws in these measures, we will review reasons for concern and suggestions for improving the validity and reliability of these measures. The ultimate purpose of this study is to challenge theorists and methodologists to develop new frameworks that will be more sensitive to complicated R/E identities while also being plausible for those using administrative data sets. The following section acts as the guiding framework for this review through summarizing a broad overview of

contemporary perspectives on the measurement and analysis of quantitative data on R/E. We juxtapose the overall approaches of those who study critical race theory and their views on quantitative R/E measurement with traditional quantitative researchers. We then describe how we utilize these viewpoints as a conceptual framework guiding our review followed by a description of our methods and results.

IMPORTANT CHALLENGES IN MEASUREMENT AND ANALYSIS OF RACE/ETHNICITY

Both individual R/E identities and R/E categories are constantly shifting and culture dependent (Denton & Deane, 2010; Liebler et al., 2017; Mihoko Doyle & Kao, 2007). Those who collect data on R/E grapple with questions like whether to ask about a person's color (e.g., Black) or someone's ethnic background (e.g., African American; Davis et al., 2012). There is also considerable heterogeneity in how administrative data sets account for those with multiple racial identities. Measurement and missing data issues create challenges in analyzing R/E data. Below, we review three perspectives on R/E measurement and analysis that help inform our subsequent systematic review: (1) a critical view of quantitative measurement and analysis of R/E most notably from the critical race theory community, (2) the perspective of quantitative researchers who engage in research on R/E, and (3) contemporary scholarship seeking to combine critical race theory with quantitative methods, specifically the developments around a QuantCrit framework. We recognize that each perspective includes several additional viewpoints and epistemologies. As we outline these perspectives to provide background for the larger literature review, we focus on providing a generalized overview.

Critical Race Theory and Measurement/Analysis of Race/Ethnicity

Many scholars have written about the inherent flaws in using quantitative methods to research R/E (e.g., Zuberi, 2001). The most prominent and historical argument traces back to the roots of the development of statistics itself. Francis Galton is known as one of the most influential statisticians in the modern era having invented some of the core tools of quantitative analysis including correlations. A half cousin to Charles Darwin, Galton sought to extend the theory of evolution into modern human reproduction by using quantitative analysis (Roberts, 2011). Galton was the founder of the eugenics movement and justified the movement by utilizing statistics to create the illusion that science backed up its tenets. To Galton, the measurability of race was for nefarious purposes—to prove his hypothesis that some races were superior to others (Covarrubias & Vélez, 2013; Sablan, 2019; Zuberi, 2001). The destruction that was caused by the eugenics movement, including genocide, mass sterilization, and pseudo-scientific based subjugation, and its continued legacy today in the White nationalism movement cannot be denied. In education, eugenics was influential in the creation of many common policies that persist today, including tracking and test score-based college admissions (Stoskopf, 1999; Winfield, 2007)—both policies that evidence

indicates continue to privilege White students over students of color (e.g., Dixon-Román, 2017; Grissom & Redding, 2015; Kobrin et al., 2007; Santelices & Wilson, 2010). Responding to the popularity of eugenics-based arguments, some of the earliest writing from W. E. B. DuBois pointed out that statistical arguments of racial inferiority ignored significant heterogeneity within Black populations as well as systemic racism (DuBois, 1899). While DuBois himself utilized quantitative data to create some of the first data visualizations (see Battle-Baptiste & Rusert, 2018), in *White Logic: White Methods*, Zuberi and Bonilla-Silva (2008) argue that the eugenics-based thinking that underlies all statistical analyses was infused within the academy such that physical and social sciences themselves have aided in the continuation of racial stratification as both scientifically legitimate and socially acceptable (Zuberi & Bonilla-Silva, 2008). Even though eugenics itself was disavowed decades ago within the academy, it remains difficult to utilize the statistical tools created by eugenicists to study R/E in ways that do not lead to perpetuating that inequality.

Scientists have now come to the consensus that race is not biological; it is socially constructed (Covarrubias & Vélez, 2013; N. M. Garcia et al., 2018; Gillborn et al., 2018; Ladson-Billings, 2012). R/E has salience for individuals only to the extent that they can self-identify with a particular R/E identity or that others categorize them with one. In addition, R/E self-identification is variable, depending on a variety of elements within the context, such as macropolitical environment or age of the individual (e.g., Liebler et al., 2017; Mihoko Doyle & Kao, 2007; Roberts, 2011). Therefore, the extent to which those who manage administrative data and analyze these data conceptualize the socially constructed and mutable nature of R/E and how it conflicts with how the individuals in the data self-identify determines how useful the data actually are.

Critical race theory in education positions structural inequality, racism, and White supremacy as inherent aspects of educational system/outcomes, acting as a framework for conceptualizing research and interpreting findings on R/E inequality (Ladson-Billings, 1998; Ladson-Billings & Tate, 1995). Stemming from the eugenics-based arguments, critical race theorists have identified three challenges to the use of quantitative measurement and analysis of R/E: claims of neutrality/objectivity, lack of discussion/recognition of power and structural aspects of racism, and White dominance in the academy. One of the central tenets of critical race theory is that any study of R/E is subjective and context dependent. This directly contradicts the argument that quantitative research is neutral, objective, and generalizable (Carbado & Roithmayr, 2014; N. M. Garcia et al., 2018; Gillborn et al., 2018; Sablan, 2019). Critical race theorists argue that quantitative research cannot be as neutral and objective as is claimed because quantitative researchers make many decisions about measurement and analysis that remain as artifacts in their results (Covarrubias & Vélez, 2013; Zuberi & Bonilla-Silva, 2008). Quantitative research also tends to focus on R/E as an individual experience. Critical race theorists recognize that racism is also structural, organizational, and institutional in addition to individual (Carbado &

Roithmayr, 2014; N. M. Garcia et al., 2018; Gillborn et al., 2018; Sablan, 2019; Tatum, 2017). Traditional quantitative researchers tend to ignore these more structural elements, focusing more on aggregating individual trends using unclear definitions of R/E.

Quantitative Perspectives on Measurement and Analysis of Race/Ethnicity

There has been little engagement between critical race theory and quantitative scholarship (Covarrubias & Vélez, 2013; Sablan, 2019). As Ladson-Billings (2012) wrote, education researchers have typically utilized R/E with a lack of attention or understanding that the categories they utilize are superficial and constructed using naïve understandings of class and race that are then imbued with deficit-oriented markers of inferiority and superiority. While some quantitative researchers raise concerns about valid and reliable measures of R/E (which we will review in this synthesis), the most prominent concerns have to do with the statistical properties of power, precision, and parsimony. These properties are all separate and intricately linked when using frequentist statistics. While none of these concepts determine the validity of a model's results, they are all important aspects of modeling decisions that can affect interpretation. All three are linked in some way to the sample size used in a study. Within the broader U.S. population (excluding studies of subsamples of students), the more complexity the researcher allows for the R/E identity of the sample, the lower the sample size. Therefore, navigating issues with power, precision, and parsimony can create barriers in the minds of some quantitative researchers to allowing additional complexity within R/E measures.

Negotiating the Two Perspectives: QuantCrit and Challenges to Quantitative Orthodoxy

A community of scholars has been purposefully attempting to combine critical race theory with quantitative research methods, calling this methodology QuantCrit (Sablan, 2019). The goal of these efforts is to create a space for quantitative research that engages with critical race theory authentically. To do so, QuantCrit scholars start by recognizing the flaws in quantitative research that quantitative researchers do not typically address (Covarrubias & Vélez, 2013; N. M. Garcia et al., 2018; Gillborn et al., 2018; Sablan, 2019). First, as noted above, R/E is socially constructed, and quantitative research needs to recognize that race is not biologically determined and static (Covarrubias & Vélez, 2013; N. M. Garcia et al., 2018; Gillborn et al., 2018). Second, QuantCrit work needs to differentiate itself from other quantitative research by taking a clearly subjective stance; recognizing that the analyst cannot be separated from the analysis is an essential component of QuantCrit (Covarrubias & Vélez, 2013). QuantCrit scholars also suggest that work should recognize the structural elements of racism (Carbado & Roithmayr, 2014; N. M. Garcia et al., 2018; Gillborn et al., 2018; Sablan, 2019). This work also needs to recognize that research and training for quantitative analysis almost exclusively occurs within White spaces, and this

will be reflected in the analysis in some conscious or unconscious way (Covarrubias & Vélez, 2013; Zuberi & Bonilla-Silva, 2008). Finally, QuantCrit needs to take an assets-oriented perspective instead of a deficit-oriented perspective (Sablan, 2019). Through either directly conducting research that avoids these pitfalls and/or being cognizant of these challenges, many are beginning to conduct research on R/E from a critical race theory perspective using quantitative methods.

HOW THESE PERSPECTIVES INFORM THIS SYNTHESIS

This systematic review recognizes the perspectives of critical race theorists, including founders of the QuantCrit framework, and other quantitative researchers. The measurement of R/E has typically lacked the degree of validity to which all parties would aspire. For this review, we seek literature that attempts to improve the validity of measurement and analysis of R/E with a particular focus on insights that are applicable to administrative data systems. We do not explicitly focus on applying QuantCrit or critical race theory to this study, we seek to be more informed by multiple perspectives on the validity and reliability of measures of R/E with a focus on administrative data.

As is recognized by QuantCrit scholars, estimates from administrative systems like the U.S. Census can help define racial inequity (Sablan, 2019). We focus on administrative data for its potential to influence policymaking and address inequality, and our review includes research that pertains directly to administrative data collection. Unlike sampling-based survey data sets, administrative data have to be collected in an efficient manner and cannot rely on weighting to account for small samples of certain R/E groups. At the same time, deciding on measures is challenging because often administrative data sets are compared with each other. How to collect and measure R/E in administrative data is a distinct challenge from analyzing survey data, especially when comparing across systems that use different R/E measures. In this review, we focus on studies that address one or both of the following questions:

1. How can scholars measure R/E in administrative data sets in the United States that takes into account how individuals self-identify, multiple identities, and shifting identities?
2. How can scholars include R/E in analyses of administrative data sets in the United States that are able to take into account multiple identities, shifting identities, and small R/E subgroups?

METHOD

We conducted searches of the Google Scholar and ProQuest databases and supplemented these searches by reviewing references lists of the resulting articles. For both searches, we used the following search terms: “race OR ethnic OR ethnicity OR racial” and “quantitative OR administrative OR classification OR measure OR measuring OR ‘secondary data.’” The use of administrative databases in

education dramatically increased in the past 15 years due to legislative pressure and incentives for states to maintain these databases as well as growing technological capacity to house and analyze these data. To reflect this shift, we bound our search to only include studies from 2001 to 2019 (the passage of the No Child Left Behind Act; McGuinn, 2015).² We focused on peer-reviewed studies for this synthesis because we are not as concerned with publication bias since the studies in this review do not rely on significant effects to increase chances of publication. This restriction was included in the ProQuest search (this could not be included in the Google Scholar search). Finally, we restricted our search to only include articles written in English.

We used three different phases of review to arrive at the final set of articles for the current study. In the first phase, we conducted the search outlined above in both ProQuest and Google Scholar. The ProQuest search resulted in 808 articles and the Google Scholar search resulted in 1,085 articles for a total of 1,893. The lead author compiled these searches into a single document, which included the article's title, authors, publication date, and abstract, if available. A group of four researchers jointly reviewed a small set of the articles (30) to determine which articles needed to be excluded based on the exclusion criteria (whether the article included discussion of quantitative data measurement and classification of R/E categories, was published in a peer-reviewed journal, and was written in English). Then the four researchers met and discussed their selections to create a shared understanding of which articles should be excluded. Once consensus was met, the four researchers split all of the compiled list of results into equal sections and reviewed the articles for exclusion. This resulted in 1,714 articles being excluded. These articles were excluded for the following reasons: duplicate of another article already included (190), not being written in English (16), no discussion of quantitative data measurement (220), no classification of R/E (1,224), and not being peer reviewed (283). These numbers do not sum to 1,708 as several articles fit into multiple categories for exclusion. The first author then reviewed this final list for any improperly excluded articles. This added back 18 articles.

In the second phase of the review, the resulting 197 articles were reviewed by the authors for the additional exclusion criteria of articles that do not include the U.S. context (58), focus on racial identity formation (specifically the psychometric properties and usefulness of Multigroup Ethnic Identity Measure), or utilize technology incompatible with administrative data collection (e.g., facial recognition technology; 15). This round of the review resulted in 73 articles being excluded and 124 articles included in the analytical set.

In the third, and final, phase of the review process, the two authors and one trained graduate assistant split the list of all of the final articles and read each article. During the review of the full articles, we again evaluated articles based on the exclusion restrictions while also noting additional literature that the articles highlighted as critical to R/E classification and administrative data analysis. Through this process,

we added an additional 21 articles to the literature review. At this phase, an additional 90 articles were excluded. We excluded these articles for the following reasons: no inclusion of quantitative data measurement (12),³ no discussion of the classification of R/E (15),⁴ no U.S. context (24), the discussion of R/E classification is not applicable to administrative databases (36),⁵ and the article was not peer reviewed (12). These numbers do not sum to 90 as several articles fit into multiple categories for exclusion. Therefore, at the end of the final phase of review, we included 55 articles in the current study.

During the third phase of review, the authors wrote analytical summaries of each included article. Once this was completed, the first author reviewed all the summaries and created categories for the emergent themes across the articles. The second author then reviewed this analysis, refined the category definitions, and reviewed the category classification of all articles. The first author then reviewed the second author's revisions and incorporated them into the final analysis.

RESULTS

Based on our review of the literature, we found the overarching themes of (1) measuring R/E, (2) missing R/E data, and (3) analysis including R/E data. We discuss each theme along with key challenges to the use of R/E measures.

Measuring Race/Ethnicity

One clear pattern across the 55 studies included in this research synthesis was the heterogeneity in approaches to measuring R/E. Between changes over time, differences across populations, and methodological choices, measuring R/E was conceptualized in dozens of ways across the 55 studies. For instance, see Table 1 for an illustration of how R/E measures have changed over time in one data set (N. M. Garcia & Mayorga, 2018). In this section, we review several themes in measuring R/E, including universal measures, federal agency guidance, pan-ethnicity measurement challenges, reliability and validity of measures, and considering alternative measurement approaches.

Universal Measures of Race/Ethnicity

Several studies advocated for a universal measurement system for R/E (Buescher et al., 2005; Idossa et al., 2018; Mays et al., 2003; Moscou et al., 2003). Among the suggestions, successful universal R/E measurement system would rely on self-reporting, include comprehensive options for ethnic identification, and have clearly defined R/E terms (Idossa et al., 2018; Moscou et al., 2003). Mays et al. (2003) specifically suggested that the federal government develops a universal R/E taxonomy that has consistent classification categories that are mutually exclusive, includes categories that are consistent with how individuals think of themselves, and facilitates reliable responses from individuals and valid analytical methods (Mays et al., 2003). When

TABLE 1
The Freshmen Survey: Racial/Ethnic Survey Question 1965–2014

Year	Question	Racial/Ethnic Category
1965–1968	What is your racial background? (Circle one/mark one)	<ul style="list-style-type: none"> • Caucasian • Negro • American Indian • Oriental • Other
1969	What is your racial background? (Mark one)	<ul style="list-style-type: none"> • White/Caucasian • Black/Negro/Afro-American • American Indian • Oriental
1970	Are you: (Mark one)	<ul style="list-style-type: none"> • White/Caucasian • Black/Negro/Afro-American • American Indian • Oriental
1971–1975	Are you: (Mark all that apply)	<ul style="list-style-type: none"> • White/Caucasian • Black/Negro/Afro-American • American Indian • Oriental • Mexican American/Chicano • Puerto Rican American • Other
1976–1989	Are you: (Mark all that apply)	<ul style="list-style-type: none"> • Oriental • White/Caucasian • Black/Negro/Afro-American • American Indian • Asian American/Oriental • Mexican American/Chicano • Puerto Rican American • Other
1990–1992	Are you: (Mark all that apply)	<ul style="list-style-type: none"> • White/Caucasian • Black/African American • American Indian • Asian American/Oriental • Mexican American/Chicano • Puerto Rican American • Other • Other Latino

(continued)

TABLE 1 (CONTINUED)

Year	Question	Racial/Ethnic Category
1993–1996	Are you: (Mark all that apply)	<ul style="list-style-type: none"> • White/Caucasian • Black/African American • American Indian • Asian American/Asian • White/Caucasian • Black/African American • American Indian • Mexican American/Chicano • Puerto Rican • Other Latino • Chinese American/Chinese • White/Caucasian • Black/African American • American Indian • Asian American/Asian • White/Caucasian • Black/African American • American Indian/Alaska Native • Native Hawaiian/Pacific Islander • Asian American/Asian
1997	Are you: (Mark all that apply)	<ul style="list-style-type: none"> • Mexican American/Chicano • Puerto Rican • Other Latino • Other • Filipino American/Filipino • Japanese American/Japanese • Korean American/Korean • Southeast Asian (Vietnamese, Laotian, Cambodian, etc.) • Other Asian American/Asian • Other • Mexican American/Chicano • Puerto Rican • Other Latino • Other • Mexican American/Chicano • Puerto Rican • Other Latino • Other
1998–2000	Are you: (Mark all that apply)	<ul style="list-style-type: none"> • Mexican American/Chicano • Puerto Rican • Other Latino • Other • Mexican American/Chicano • Puerto Rican • Other Latino • Other • Mexican American/Chicano • Puerto Rican • Other Latino • Other
2001–2014	Are you: (Mark all that apply)	<ul style="list-style-type: none"> • Mexican American/Chicano • Puerto Rican • Other Latino • Other • Mexican American/Chicano • Puerto Rican • Other Latino • Other • Mexican American/Chicano • Puerto Rican • Other Latino • Other

Note. Reprinted from N. M. Garcia and Mayorga (2018). Used with permission.

discussing a national model birth certificate that includes R/E classification, Buescher et al. (2005) cautioned that while standard categories increase reliability and facilitates comparisons across states, standardization does not mean R/E will have salience when people do not understand the concept of R/E such that “a broadly defined racial group is at best a crude marker . . . certainly not a risk factor or cause” (Buescher et al., 2005, p. 397). While all administrative data systems in the United States do not currently use a universal R/E classification system (e.g., different states can determine the definition of their R/E measures), the closest measure we have are those utilized by the U.S. Census Bureau as well as other federal agencies.

U.S. Census Bureau and Federal Agency Guidance on Measuring Race/Ethnicity

The U.S. Census has collected data on R/E since 1790 with significant changes in measurement and categories over time (Kilty, 2004; Mays et al., 2003; Rodriguez, 2000). The modern R/E Census categories were based on the Office of Management & Budget Directive 15 from 1977 that specified race should be reported in four mutually exclusive categories: White, Black, American Indian or Alaskan Native, and Asian or Pacific Islander. Ethnicity was categorized as Hispanic or not of Hispanic origin (Mays et al., 2003). Prior to this directive, there was no Census question on Hispanic origin with those who identified as being from Latin America primarily seen as White (Idossa et al., 2018; Mora, 2014). See Table 2 for a visual representation of how R/E categories have changed over time on the U.S. Census.

Race/ethnicity measurement on modern Census forms. Studies on the 1990, 2000, and 2010 Census focused on several specific measurement choices that had greatly affected R/E counts. For all of these Census administrations, two separate questions asked about race and Spanish/Hispanic origin. Those of Spanish/Hispanic origin in the 1990 Census were the largest group to mark “Other” for their race (Mays et al., 2003), with 97% of those who selected “some other race” in 2000 identifying as Spanish/Hispanic/Latino and 42% of Spanish/Hispanic/Latino respondents selecting “some other race” (Campbell & Rogalin, 2006).

Other changes in the 2000 Census were in direct response to advocacy groups lobbying for recognition of specific R/E groups. The biggest change allowed respondents for the first time to select more than one R/E category (Aspinall, 2003; Idossa et al., 2018; Mays et al., 2003; Prewitt, 2005, 2018). While many advocated for this change, civil rights advocates, including the National Association for the Advancement of Colored People (NAACP), were concerned that allowing multi-R/E selection would diminish the size of discrete minority populations (Prewitt, 2005, 2018). While R/E count estimates did not markedly change, almost 2 million people selected Black and another R/E category (Campbell, 2007).

Persistent challenges for race/ethnicity measurement on the Census. After no substantive changes were made to R/E measurement in the 2010 Census, many persistent

TABLE 2
Changes in U.S. Census Race Categories From 1850 to 2000 (No Changes in 2010)

Year	White	Black	Native People	Chinese	Japanese	Other Asian or Pacific Islander	Other
1850		Black, mulatto					
1860		Black, mulatto	Indian				
1870	White	Black, mulatto	Indian	Chinese			
1880	White	Black, mulatto	Indian	Chinese			
1890	White	Black, mulatto, quadroon, octroon	Indian	Chinese	Japanese		
1900	White	Black	Indian	Chinese	Japanese		Other (+ write in)
1910	White	Black, mulatto	Indian	Chinese	Japanese		Other (+ write in)
1920	White	Black, mulatto	Indian	Chinese	Japanese	Filipino, Hindu, Korean	Other races, spell out in full
1930 ^a	White	Negro	Indian	Chinese	Japanese	Filipino, Hindu, Korean	(Other races, spell out in full)
1940	White	Negro	Indian	Chinese	Japanese	Filipino, Hindu, Korean	(Other races, spell out in full)
1950	White	Negro	American Indian	Chinese	Japanese	Filipino	(Other races, spell out)
1960	White	Negro	American Indian	Chinese	Japanese	Filipino, Hawaiian, part Hawaiian, etc.	Other (print race)
1970	White	Negro or Black	Indian (American)	Chinese	Japanese	Filipino, Hawaiian, Korean	Other (specify)
1980	White	Black or Negro	Indian (American), Eskimo, Aleut	Chinese	Japanese	Filipino, Korean, Vietnamese, Asian Indian, Hawaiian, Guamanian, Samoan	Other (specify)
1990	White	Black or Negro	Indian (American), Eskimo, Aleut	Chinese	Japanese	Filipino, Hawaiian, Korean, Vietnamese, Asian Indian, Samoan, Guamanian, other Asian or Pacific Islander	Other race
2000	White	Black, African American, or Negro	American Indian or Alaska Native	Chinese	Japanese	Filipino, Hawaiian, Korean, Vietnamese, Asian Indian, Samoan, Guamanian, other Asian or Pacific Islander	Some other race (specify)

Note. See notes on Table 1 in Mays et al. (2003) for more information. Re-created from Mays et al. (2003) with updates.

^aIn the 1930 Census, Mexican was included as a racial category.

challenges and methodological issues remain that could be remedied in a future Census. Chief among these issues is that asking separately about Hispanic origin and race has consistently led to an inflated “other” race category made up predominately of those of Hispanic origin. Federal surveys in general asked these questions separately to not conflate common ancestry, language, and culture (i.e., ethnicity) with common physical/phenotype characteristics (i.e., race; Campbell & Rogalin, 2006; Eisenhower et al., 2014; Mays et al., 2003). However, those who respond to the Census form and other government surveys have been shown to not make the same distinction (Campbell & Rogalin, 2006; Eisenhower et al., 2014). Despite being part of the Census form for several decades, empirical evidence has shown that this distinction between Hispanic origin and R/E is not resulting in valid data differentiating those of Hispanic origin by R/E.

Two specific recommendations for the 2020 Census were made in the literature reviewed. First, combine the separate Hispanic question with the overall question on race for a single R/E question. Respondents would still be able to mark more than one category and write-in options would still be available. The other recommendation was to add “Middle Eastern, North African” as an option on the R/E question. These groups have traditionally been considered White by federal data sets (Prewitt, 2018). Since Prewitt (2018) published this article, the U.S. Census Bureau has submitted a request to keep two separate questions on Hispanic origin and race. The request also specified there would be no separate Middle Eastern, North African category, instead keeping these categories explicitly as part of the White option (U.S. Census Bureau, Commerce Department, 2018).

Despite the importance of the U.S. Census as a model that other administrative data sets can use, scholars have considered different ways to measure R/E that address these persistent challenges. We review many of these advances below but first highlight the challenges of Hispanic pan-ethnicity.

The Hispanic Pan-Ethnicity Presents Specific Measurement Challenges

As discussed above, asking separately about Hispanic origin and race has not successfully led to differentiating ethnicity from race with the plurality of Hispanic respondents choosing “other” as their race (Eisenhower et al., 2014; Haney López, 2005; Hitlin et al., 2007). If given an open response option, Hispanic respondents tended to write in their country/region of origin (Idossa et al., 2018; Landale & Oropesa, 2002), or, on the Census in particular, they wrote in “Latino” (Haney López, 2005). Studies have suggested that more granular measures of R/E would resolve this issue (Hitlin et al., 2007; Prewitt, 2018). These measures could include country of origin, language, religion, migrant status, nationality, skin color, geographic region, and recency of immigration (Aspinall, 2009; DiPietro & Bursik, 2012; J. A. Garcia, 2017; Haney López, 2005; Idossa et al., 2018; Williams & Husk, 2013). Including all of these various alternative ethnicity measures has proven to be politically sensitive at the federal level with concerns about questions related to

nativity and citizenship potentially undermining data accuracy and privacy. At the same time, including more specific questions about language and country of origin likely improve reliability, especially for more recent immigrants, as pan-ethnic identity identification has been shown to be less reliable over time especially for adolescents (Feliciano & Rumbaut, 2018, 2019). In addition, while including measures of color (e.g., hair color, facial features, and skin tone) could be illuminating for differentiating inequality within the Hispanic ethnicity, Haney López (2005) argued that the Census and other administrative data collection efforts are unlikely to include these measures for political reasons.

Reliability and Validity of Race/Ethnicity Measures

Across any measure, it is important to be aware of and explicitly examine the reliability and validity of said measure. From a historical perspective, defining R/E has been entirely reliant on cultural norms and time dependent (Roberts, 2011). Defining what it means to be a valid and reliable measure of R/E is in and of itself a difficult enterprise. The included studies focused on assessments of reliability or validity that gauged how frequently these measures might change over time or be recorded inaccurately. We conceptualize measurement reliability of R/E as how consistently individuals report R/E when asked extremely similar questions about their R/E over time. Measurement validity of R/E measures refers to the accuracy of the R/E measures when comparing R/E measures across multiple sources of data, a form of concurrent validity.

Reliability. Issues of reliability tend to focus on those with multi-R/E identities. While allowing individuals to select more than one R/E hypothetically increases the validity of responses (i.e., more accurately represent how people see themselves), many have expressed the concern that this multi-tick option decreases reliability (Aspinall, 2009; Prewitt, 2018). For example, 40% of individuals who checked multiple boxes on the R/E question on the 2000 Census identified as monoracial in a follow-up survey 1 year later (Prewitt, 2018). Another concern is that many multiracial individuals will be more likely to exclusively check the “White” box as they assimilate (Prewitt, 2018). Open-response R/E measures that are often used in conjunction with multi-tick boxes have been shown to be particularly unreliable (Aspinall, 2001). These reliability challenges often stem from the lack of agreement on what it means to claim certain R/E for those whose R/E identification is culturally and temporally dependent. For instance, Roberts (2011) wrote about how President Barack Obama identified solely as Black on the 2010 Census. At other points in his life, he might have identified as White and Black and in other time periods, he would have been identified by Census data collectors as “mulatto.”

Several studies examined R/E responses over time of multiracial individuals to examine reliability. For instance, Harris and Sim (2002) compared R/E measures that were collected at school with those collected at home. Twice as many students

self-identified as multiracial at school than at home. This change was mostly due to more students identifying as White–American Indian at school than at home. Less than 2% of students consistently identified themselves as multiracial (Harris & Sim, 2002). Another study examined changes in multiracial identification over time (across 6 years) finding that 6% of adolescents changed their R/E, almost half of which went from being monoracial to being multiracial (Hitlin et al., 2006). Just as with the previous study, self-classification change was often due to changing American Indian self-categorization (Harris & Sim, 2002; Hitlin et al., 2006).

Two studies also compared how R/E identification could differ for children based on the R/E of their parents. Both of these studies found that having parents of different races did not necessarily mean that they identified their children as multiracial, and having parents of the same race did not necessarily mean the child would not be multiracial. Multiracial identification was more common when one parent was White or American Indian and the other parent was not (Bratter, 2007). In families with one Black and one White parent, about half identified their child as Black-White, a quarter as Black, a tenth as White, and a tenth as “other” (Roth, 2005).

Other studies examined reliability of R/E more broadly outside of just multi-R/E identities. For instance, Feliciano and Rumbaut (2018, 2019) examined how ethnic self-identity labels change from adolescence to early adulthood among children of immigrants. They found about half of their sample kept the same ethnic identity over time with changes less likely to occur during adulthood. Using pan-ethnic labels or identifying as “American” was common during adolescence but much less common in adulthood (Feliciano & Rumbaut, 2018, 2019).

Craemer (2010) examined if being reminded of genetic or ancestral information would induce changes in R/E self-identification within a short time (5–90 minutes). He found that about 3% of the sample made short-term self-classification changes with American Indian/Alaskan Native, “other,” and multiracial categories tending to lose members while Black, Hispanic, and Asian categories tended to gain members (Craemer, 2010). It is possible that R/E self-identification will lose reliability over time as ancestral genetic research gains popularity and specificity.⁶

Validity. Studies assessing the concurrent validity of R/E measures tended to use data from health records and compared records either across health systems or randomly selected individuals to be surveyed about their R/E to compare with their health records. Overall match rates across data sources tended to be around 60% to 70% when restricted to those with complete information (Eisenhower et al., 2014; Kressin et al., 2003; Moscou et al., 2003; Smith et al., 2010). Accuracy tended to be lower for smaller R/E groups. For instance, Kressin et al. (2003) found that agreement rates were 60% for African American, Hispanic, and White but 15% for American Indians. Smith et al. (2010) found accuracy among White records to be 89% and 18% among American Indian/Alaskan Native. Several studies also found that

rates of missingness differed across populations with Hispanic individuals more likely to be missing R/E information in one of the sources of data (Eisenhower et al., 2014; Kressin et al., 2003; Maizlish & Herrera, 2006; Smith et al., 2010). Overall, the agreement on R/E across data sets left much to be desired. Authors of these studies encouraged collecting R/E through self-report whenever possible (since this is health care data, on provider visit or hospitalization), linking records to birth certificates, creating measures with more granularity for Hispanic individuals, and uniform data collection procedures (Eisenhower et al., 2014; Kressin et al., 2003; Moscou et al., 2003; Smith et al., 2010).

Considering Alternative Measurement Approaches

Defining race/ethnicity categories. While many have lamented the limitations inherent in common measures of R/E, especially those utilized by the U.S. Census Bureau, the difficult task remains of how to improve the options available for measuring R/E. When deciding on R/E categories, there was an inherent trade-off between statistical reliability and validity (Aspinall, 2001; Buescher et al., 2005; Eisenhower et al., 2014; Williams & Husk, 2013). For instance, the White category was very reliable but can mask important variation and the disadvantages of certain ethnic groups like those from the Middle East in the United States (Williams & Husk, 2013). At the same time, including measures that were increasingly multidimensional might not be practical due to respondent and administrative burden (Aspinall, 2001; Eisenhower et al., 2014). Likely a solution lies in disaggregating some broad categories and by being mindful of the flaws inherent in the chosen approach (Eisenhower et al., 2014; N. M. Garcia & Mayorga, 2018).

Several studies include recommendations when defining R/E categories. First, it is important to think about whether R/E categories should be defined by color (e.g., Black) or by “racial” group (e.g., African American), and this decision can have important implications for how people self-identify (Davis et al., 2012; Eisenhower et al., 2014; Roth, 2010). Eisenhower et al. (2014) suggested using color, while Roth (2010) noted that this decision should be based on what is intended to be measured. Roth (2010) created a schematic framework where racial self-identification can be based on subjective self-identification, the race you tell others, the race others believe you to be, among other options. While R/E measures often intend to separate color from ethnicity, it was difficult to proxy racial differences for those who solely identify with their ethnicity—namely, for those of Hispanic origin (Eisenhower et al., 2014; Roth, 2010). Second, selecting R/E categories can begin by determining the universe of all R/E categories such that the list represents how the majority of individuals would self-classify if asked their R/E. This process needs to be done sensitively as some differentiation lacks salience like differentiating the White population in the United States by European country of origin (Marquardt & Herrera, 2015). Any list of categories utilized in U.S. administrative data sets will likely focus on race rather than on ethnicity since ethnicity is seen as diverting attention away from issues of

structural racism and power (Aspinall, 2001). Also, there are individuals throughout the United States who may not know their exact ethnic origins (e.g., Black descendants of slaves).

Most identify/best represents items. One of the causes of unreliable R/E measurement (different responses being selected in different data sets) was allowing respondents to select multiple categories. A proposed measurement solution in the literature to help ameliorate this reliability issue was to follow up on R/E questions with an additional item asking those with multiple responses the R/E they most identify with, or which R/E best represents their identity (Campbell & Rogalin, 2006; Mays et al., 2003; Parker et al., 2004; Williams & Husk, 2013). An item gauging the strength or importance of each dimension of R/E of that person allows for more reliable data since the respondent is less likely to change their most salient identity over time or across data sets. This type of item can also ease analysis (discussed more below) though it does not help when individuals do not identify with one dimension of their R/E more than others (e.g., if their primary identity is multiracial).

Phenotype. R/E is often assumed to measure some sort of common experience and is useful for identifying disparities and discrimination. However, there is great variation in experiences and discrimination within R/E categories. Prior work has found phenotype to be highly influential in educational outcomes with systemic disparities within racial group by skin tone (see Monroe, 2013). Several studies suggested measures for skin tone or phenotype as a way to better measure these within-R/E disparities (Foy et al., 2017; J. A. Garcia, 2017; J. A. Garcia et al., 2015; Roth, 2010). For instance, J. A. Garcia et al. (2015) utilized the question, “We are interested in how you would describe your appearance. How would you describe your skin color with 1 being very light to 5 being very dark or somewhere in between?” (p. 359). While items such as this one have been utilized, more research on the reliability and validity of self-reported phenotype items would be necessary for inclusion in broader administrative data collection.

It is important to point out that differences by phenotype are distinct from differences between internal (subjective self-identification) and observed (the race others assume you to be) R/E (Roth, 2010; Vargas & Kingsbury, 2016). Racial identity contestation refers to when one identifies as one R/E but is perceived by most others as a different R/E. About 6% to 10% of adults experienced racial identity contestation, and it is most common for American Indians, although could become more common over time with increasing rates of interracial marriage (Vargas & Kingsbury, 2016). While it remains rare for administrative data sets to include phenotype, it might be an area to consider especially if the Hispanic option becomes part of the full list of R/E categories since separate R/E questions were designed to assess whether an individual identified as Hispanic separate from their race/phenotype/physical appearance.

Missing Data: An Analysis and Measurement Concern

When quantitative education researchers utilize administrative data sets, it is a general expectation that R/E data will, at the very least, be included as covariates regardless of the analytical design. One of the key problems that can arise, however, is missing data on R/E. Those analyzing administrative data sets, especially in education, likely treat missing data on R/E like any other missing data problem. This would lead to a set of common solutions to missing data, including complete case analysis (i.e., dropping observations with missing R/E), an indicator variable for “missing R/E” to account for missing observations, and multiple imputation (see Thompson et al., 2018). At the same time, a robust literature base in public health, epidemiology, linguistics, and other similarly situated fields has utilized other personal information from administrative data files often in conjunction with advanced statistical methods to address missing R/E information (e.g., Adjaye-Gbewonyo et al., 2014; Fremont et al., 2016; Kilty, 2004; Mateos, 2007). These methods use lists of common first names and/or surnames sometimes in conjunction with geocoded address block or tract-level information to assign a probability of a certain R/E or even a specific R/E.

The original name-based R/E classification systems utilized surnames to assign individuals to an R/E (Kilty, 2004; Mateos, 2007). The U.S. Census Bureau has maintained a Spanish/Hispanic surname list since the mid-20th century to identify Hispanic individuals (Fiscella & Fremont, 2006; Kilty, 2004; Voicu, 2018). Researchers have also developed surname lists to identify Asian surnames, in general, as well as Chinese, Indian, Japanese, Korean, Filipino, and Vietnamese American surnames, specifically (Fiscella & Fremont, 2006; Mateos, 2007). Surname lists in the United States are limited to identifying those that could potentially be Hispanic or Asian with little to no utility in identifying other R/E categories. In other countries, surname lists have been utilized to identify religious groups and those of Middle Eastern descent (Mateos, 2007). Others have created first-name based lists to identify R/E using first name and surname, finding that first names might more accurately identify White individuals than surnames (Tzioumis, 2018). As reference name lists continue to be created, Mateos (2007) cautions researchers to make sure that these lists were based on a large enough population to make valid inferences and to be aware of temporal differences, regional differences, differences in average ratio of people per surname, history of name adoption, and surnames reflecting only patrilineal descent.

The theory behind identifying geocoded addresses was that block groups or neighborhoods tend to be racially homogeneous for certain R/E groups, particularly in more segregated regions of the United States. The method proxied the probability someone is a certain R/E by examining the R/E composition of those who live in close proximity to the individual. This method has been found to be most accurate when identifying Black and White individuals (Elliott et al., 2008; Fiscella & Fremont, 2006) but inaccurate for identifying American Indian/Alaskan Native and/

or multiracial individuals (Fremont et al., 2016; Voicu, 2018). Using geocoded addresses lacks accuracy in regions of the country with lower residential segregation and with non-White and non-Black populations who tended to live in more integrated neighborhoods.

The majority of recent studies within this literature on using addresses/names to account for missing R/E assessed the accuracy of the RAND Corporation's Bayesian Improved Surname Geocoding (BISG) method. Researchers used this method to combine names and geocoded addresses to produce probabilities that an individual is Hispanic, Black, Asian, or White. The method first calculated probabilities using surnames from the Hispanic and/or Asian surname lists followed by updating those probabilities based on geocoded Census blocks. Work assessing the accuracy of BISG did so by comparing these predicted probabilities from the Bayesian model to actual self-reported R/E from administrative data sets (usually from the health field; Adjaye-Gbewonyo et al., 2014; Elliott et al., 2008; Fiscella & Fremont, 2006; Fremont et al., 2016; Grundmeier et al., 2015; Shah & Davis, 2017; Voicu, 2018). A recent article updated the BISG to incorporate an additional step of updating probabilities based on first names, which marginally improved the accuracy of identifying those whose self-reported R/E is Black (Voicu, 2018).

The accuracy and the reported utility of the BISG somewhat differed depending on study. For instance, Fremont et al. (2016) reported that the correlation between BISG estimates and self-reported race was 0.90 to 0.96 for Black, Asian American/Pacific Islanders, Hispanic, and White individuals, although the authors note that RAND discouraged using BISG probabilities to assign individual R/E. Elliott et al. (2008) stated that the average correlation between BISG probabilities and self-report was 0.70. Both studies, while reporting substantially different correlations, suggested using BISG for aggregating disparities overall, or the probabilities themselves could be utilized in regression in place of binary indicators of R/E (Elliott et al., 2008; Fremont et al., 2016). When aggregating estimates, Adjaye-Gbewonyo et al. (2014) suggested a method for determining probability cutoffs to assign R/E that takes the trade-off between inaccuracy and individuals not being assigned to an R/E into account. In their study, they suggested that when individuals have an R/E probability of at least 0.50 to 0.57, they be assigned that R/E although they caution that these cutoffs could change depending on the population. One study explicitly compared BISG with the traditional methods of handling missing data, finding that BISG-enhanced imputation significantly reduced bias compared with complete case analysis, using indicators for missing values, and multiple imputation (Grundmeier et al., 2015).

Analysis Incorporating Race/Ethnicity

Fewer articles in this synthesis critically engage with questions on how to incorporate R/E into quantitative analysis as compared with the literature on the measurement of R/E (14 articles are reviewed in this section versus the 47 included information

on measurement). Several essays and reviews pointed out a common theme in including R/E in analysis: doing so without nuance, description, or thoughtfulness. While R/E were socially constructed, studies tended to include R/E as covariates with the assumptions that these variables are independent or causal (N. M. Garcia & Mayorga, 2018; James, 2001; Lee, 2009; Ma et al., 2007). For instance, James (2001) described how race cannot cause things to occur, “Instead of merely ‘controlling’ for the difference of the aberrant ‘others,’ racial differences should be assessed and grounded in the set of historical and social circumstances that give meaning to the race concept” (James, 2001, p. 246).

Several articles encouraged researchers to be intentional about common analysis decisions like including R/E as covariates and examining effect heterogeneity by R/E (N. M. Garcia & Mayorga, 2018; James, 2001; Lee, 2009; Ma et al., 2007). Often diversity in R/E was recoded into falsely homogeneous indicators like non-White or “students of color,” and these methodological decisions should be explained and acknowledged (N. M. Garcia & Mayorga, 2018). Other suggestions included clearly stating how R/E was measured and operationalized as well as the relevance of R/E in the study and why R/E is in the models (Lee, 2009). Another suggestion was for fields to enforce consistent definitions. In a review of studies from the four highest impact medical journals, Ma et al. (2007) found a total of 116 terms for R/E categories with at least 10 different terms being utilized to describe each of the major R/E categories, including White, Black, Asian, or Hispanic. Consistent definitions and including information about measurement and operationalization are important for judging the validity and accuracy of the data (Lee, 2009; Ma et al., 2007).

Another important area of concern that Lee (2009) noted is the tendency to claim that the lack of heterogeneity by R/E was solely because of lack of power (i.e., a small sample size too small to detect an effect) or to explain heterogeneity by R/E that utilized biological or genetics-based arguments. As R/E is socially constructed, heterogeneity by R/E has social or environmental components that are rarely acknowledged within the medical field. As Lee wrote,

This biomedical and genetic focus may lead to biomedical solutions and the withdrawal of social, political, or economic approaches to easing social and economic inequalities. Furthermore, we may inadvertently accept the validity and legitimacy of a biological understanding of race. (Lee, 2009, p. 1189)

Two studies focused specifically on ideas for incorporating R/E into analysis in innovative ways. One example was drawn from the critical race theory tradition coining the term Critical Race Quantitative Intersections + Testimonios (CRQI+T). The CRQI+T framework encouraged directly incorporating qualitative information to make data more experiential and to disrupt dominant data mining techniques (Covarrubias et al., 2018). Under CRQI+T, data mining and analysis were guided by personal experiences and testimonios, and informed by qualitative perspectives. Unlike traditional quantitative research, which defines significance based on statistical calculations of the change of type I error (e.g., $p < .05$), the CRQI+T framework

defined the significance of findings based on the testimonies or qualitative perspectives (for an example, see Covarrubias et al., 2018). Mayhew and Simonoff (2015a, 2015b) addressed the common decision to make White-only the reference group within models that included a series of binary indicators for R/E. Instead of making White-only the normative category, the authors suggested what they termed an *effect coding* approach of eliminating White-only as the reference group by recoding each binary indicator of R/E to have a value of -1 for White. They argued that this approach takes an assets-driven modeling approach, while traditional White as a reference category approaches were deficits based. This approach can also be useful for incorporating complexity of R/E identity into models. Thus far, this approach has gained some traction within higher education research with fewer inroads into the K-12 education research community based on current citations.

Incorporating Complex Race/Ethnicity Measures Into Analysis

While few studies more generally considered how to analyze R/E, a broader literature has critically examined the analysis of multi-R/E measures of identity. This literature has been spurred by two interrelated, relatively new ways of measuring R/E: having multiple items to measure complexity in R/E identity and allowing respondents to select more than one R/E option (J. A. Garcia, 2017; Prewitt, 2005). These measurement choices have several important implications for analysis. First, allowing for complex self-identification often led to groups that are too small for statistically significant estimates (in frequentist statistics) in addition to privacy concerns (Marquardt & Herrera, 2015; Prewitt, 2005; Saperstein et al., 2016; Williams & Husk, 2013). For instance, having two R/E items and allowing for more than one R/E box to be checked led the 2000 Census to have up to 189 possible R/E combinations (Prewitt, 2005). Second, Marquardt and Herrera (2015) warned that having more complex measures of R/E identity did not mean that all of these identities will be salient to each individual. When analyzing data with complex R/E identities, it could be advantageous for the researcher to consider only the R/E options that were politically relevant for the study. Political relevance was determined by the context and focus of the study. For instance, this could be determined by including only ethnic groups that are represented by a political organization or groups determined to be at risk for conflict (Marquardt & Herrera, 2015). Third, added complexity has the potential to produce tension between model parsimony and validity (J. A. Garcia, 2017). The appropriate analytical techniques for capturing complex R/E identities might not be obvious, and it could be necessary to examine multicollinearity and sample sizes of small cells (J. A. Garcia, 2017; Saperstein et al., 2016).

Several studies tested specific methods for analyzing complex R/E data, including how to bridge across data sets using different techniques, how to assign single-race categories, and how to compare measures of different dimensions of R/E identity. Those analyzing complex R/E data might have a need to assign multi-R/E individuals to a single R/E for analytical reasons or to combine two data sets that utilize

different R/E measures. How this reassignment is done can have major implications for the results depending on the context (Campbell, 2007; Mays et al., 2003; Parker et al., 2004; Schenker & Parker, 2003). For instance, Campbell (2007) considered those who identified as multiracial with one racial identity being Black. While these individuals represented 0.6% of the U.S. population, if their other racial identity was a more rare identity like American Indian/Alaskan Native then reassignment can have major implications.

Reallocation methods can take into account a measure of the R/E the person most identified with (if available) or several single-race assignment options using arbitrary decision rules (e.g., choosing Hispanic ethnicity over all other R/E or assigning the rarest R/E; Mays et al., 2003; Parker et al., 2004; Schenker & Parker, 2003). Arbitrary decision rules were discouraged although a measure of which R/E the person most identifies with was often unavailable (Mays et al., 2003) and may not be an appropriate measure. One option when this measure was not available, according to the literature, is to use fractional methods for aggregating data (not for use with individual assignment) where fractions of multiracial individuals were distributed to their various R/E identities either equally or based on an algorithm. Regression models incorporating covariates can be used to create these fractions. These kinds of regression models can be relatively accurate at re-creating aggregated R/E statistics as long as the fractional distributions can vary by R/E identity (Parker et al., 2004; Schenker & Parker, 2003).⁷

As discussed above in this chapter, administrative data sets might also include multiple different measures of R/E identity. Saperstein et al. (2016) wrote about several different methods for analyzing data to account for multidimensionality of R/E, with multidimensionality being measured through self-report, phenotype, observer classification, and how people think others see them. The researcher can purposefully examine differences between the two dimensions of race by having one measure predict the other. Other options include comparing effects in models using the different dimensions of race, examining effects when all measures are included in the model simultaneously, and including a saturated model with indicators for every possible combination of identities. The analyst might use AIC (Akaike's information criteria) and BIC (Bayesian information criteria) statistics to identify which model has the most parsimonious fit to the original data (Saperstein et al., 2016).

DISCUSSION: HOW TO MEASURE AND ANALYZE RACE/ETHNICITY DATA IN ADMINISTRATIVE DATA SETS

Measurement

The reviewed studies paint a complicated picture of how to better measure and analyze R/E data from administrative data sets. A universal measure is unlikely to solve all of the measurement/analysis issues that can arise. However, creating more universal ways of measuring R/E are possible and should incorporate a few key insights from the reviewed studies. First, we clearly need to reconsider how Hispanic

origin is measured. While the U.S. Census stagnates, continuing to have two separate R/E questions, those creating measures for administrative data sets in education do not have to follow this model. Overwhelming evidence indicates that those who identify as Hispanic are unlikely to identify a separate R/E, and forcing Hispanic people to identify a separate R/E leads to issues with missing data. The literature suggests one alternative is to create one item listing possible R/E identities, including Hispanic origin, giving respondents the option of selecting more than one option. This alternative may place additional burden on the data administrators if the data must be reported to a federal or state agency with different definitions of R/E.

However, reliability is lessened by allowing respondents to check multiple boxes and giving write-in options since those with multi-R/E identities are the most likely to change their reported R/E over time especially those who are part American Indian/Alaskan Native. Collecting data over time instead of forcing individuals to have static R/E identities can help address this issue with reliability. For example, state educational administrative data sets could collect data on R/E every year to allow researchers to incorporate the fluid nature of R/E identification. This would be especially important for data on students since several reviewed studies found that multiracial and immigrant children often change their reported R/E until they reach adulthood when their reported R/E becomes more stable. Having measures of R/E over time will be especially important for those who can identify as American Indian/Alaskan Native since this identity is associated with lower reliability of responses.

Including questions that introduce complexity without confusion will be part of a more valid measurement system. For instance, measures of country of origin, language, and recency of immigration are more likely to be reliable and provide important information on R/E, particularly when focused on pan-ethnic measures of R/E.

The studies suggest a few other ways to increase the concurrent validity of R/E measures. Administrative data sets could measure R/E using self-reports whenever possible as well as supplementing R/E measurement with data from valid sources like birth certificates. Those collecting administrative data could use uniform procedures for collecting and recording this information.

When reconsidering or creating measures of R/E, there are a few concrete suggestions to consider. First, be attentive to the list of possible R/E identities. This list should represent the majority of the population, include politically relevant groups, and include heterogeneity to the extent that it will be helpful for identifying disparities. Two measures could be considered as additions. When people are allowed to select more than one R/E identity, a useful follow-up question could be to ask which of these identities best represents how they view their R/E. That said, as we mention above, there are individuals for whom multiple R/E identities equally represent their lived experiences. For these individuals, asking which identity best represents them will not mitigate researchers' issues with multiple categories being selected. For specific R/E identities, it might also be advantageous to consider self-reported phenotype to identify differences in outcomes by skin tone.

Missing Data

A robust literature primarily from public health suggests that when administrative data sets are missing R/E data on individuals, but have information on names and home addresses, the BISG method is potentially useful. However, these types of methods have several notable limitations that we urge others to carefully consider. The BISG methods are useful only for assessing the probability that an individual is Asian, Hispanic, Black, and, in some cases, White with little utility with other R/E categories. In addition, identifying these Black and White individuals relies on residential segregation, which greatly varies depending on the region. Finally, Asian and Hispanic surnames generally reflect patrilineal descent, an increasingly tenuous assumption considering growing rates of those of multi-R/E descent. The mixed evidence of the accuracy of BISG should also be considered before implementing this method.

Analysis

For analysis of R/E data, much of what the analyst can do is to better describe their measures and their analytical choices. When describing quantitative work incorporating R/E, the researchers should explain why R/E is included in a model, why the R/E measures are operationalized the way they are, and/or why they are interested in heterogeneity by race. These explanations should avoid biological or genetics-based theories instead recognizing that race is not a cause, is socially constructed and can signify common social or environmental experiences. The researcher should also be clear about how R/E was measured (e.g., self-report). Research from the medical field also suggests that education researchers consider making fieldwide definitions of R/E that should be used across studies. These standards for R/E measurement and reporting can incorporate much of the measurement advice described above.

When considering analytical choices made during modeling, researchers should attempt to integrate complex R/E identities into their models. While doing so, they should remain aware of issues of multicollinearity, small cell size, and data privacy. When these issues arise, they can consider incorporating information on R/E identities that are more salient for those respondents (if that measure is available). Researchers could also stop focusing on the binary notion of whether an estimate is statistically significant or insignificant and instead focus on the actual p value and interpret it within the context of the study (e.g., Wasserstein et al., 2019).

CONCLUSION


As administrative data sets in education continue to grow in complexity and as more researchers utilize these data sets, it will be increasingly important to attend to measures of R/E and how those measures are incorporated into analysis. While quantitative analysis of R/E brings with it a troubling history of racial subjugation, embracing quantitative methods to study and address R/E inequality can also hold


much promise in the future for understanding and lessening inequality especially with the growth of availability and analysis of big data (see Dixon-Román, 2017). Much of the work that has been done on R/E measurement and analysis has focused on medical or public health fields where administrative data have been prominent for a longer period of time. Administrators, researchers, and policy stakeholders generally use education administrative data as if each piece of information in the data is an objective fact. However, R/E are subjective, varying, and socially constructed (Covarrubias & Vélez, 2013; N. M. Garcia et al., 2018; Gillborn et al., 2018; Ladson-Billings, 2012), and educational measurement of R/E can and has been sporadically conducted. For example, Ford (2019) found state requirements that missing R/E data in K–12 administrative data sets be supplied by observer identification (as opposed to self report) in several states. However, there is no systematic policy on who can be considered the observer in this scenario or what criteria the observer should use to assign R/E values for students. These data are required for federal and state accountability and are consistently used by researchers, yet there is little clarity around the source of information on R/E measures. The current review has highlighted some of the ways that scholars can approach the use of R/E measures in educational administrative data; however, there is significant work still to be done understanding how this type of information is collected. As administrative data in education continue to grow, researchers might consider conducting studies similar to the medical and public health fields to help inform the education research base.

ACKNOWLEDGMENTS

We are grateful to Elisa Wolf and Devon Lockard for their assistance in preparing this chapter and organizing the articles for the literature synthesis.

ORCID iDS

Samantha Viano  <https://orcid.org/0000-0002-9229-3597>

Dominique J. Baker  <https://orcid.org/0000-0001-8570-9681>

NOTES

¹While DuBois offered one of the first criticisms of the use of quantitative methods to study differences between racial groups, he also created some of the earliest data visualizations utilizing quantitative data (Battle-Baptiste & Rusert, 2018).

²Google Scholar included studies with publication dates in 2019 even though the search was performed in December 2018.

³Articles excluded due to a lack of inclusion of quantitative data measurement generally focused on legal analysis and broad-based discussions of survey-based data sets.

⁴Articles excluded due to a lack of discussion about the classification of R/E did not include a focus on the operationalization or measurement of R/E.

⁵Articles excluded due to the discussion of R/E classification not being applicable to administrative databases generally included visual classification (e.g., facial recognition software), specimen collection (e.g., genetic data), or probability-based complex sample survey designs. These designs rely on sampling a portion of the population and using probability weighting

to make that sample's data generalize to the larger population. This literature review is focused on administrative data, which includes the data on the entire population of individuals and requires different statistical assumptions than survey data. Therefore, articles using survey data were excluded.

⁶We recognize that race is not a biological or genetic trait. Genetic testing, at best, reports on ancestry, not race (Roberts, 2011). However, research does show that some individuals shift their self-reported classification of R/E when provided information on genetic ancestry (Craemer, 2010). Roberts (2011) included a particularly in-depth discussion on the use and misuse of genetic ancestry tests to claim a certain R/E, including doing so as a means of claiming certain privileges based on that R/E, like tribal citizenship for American Indian genetic ancestry and Israeli citizenship for Jewish ancestry.

⁷It is important to note that this assessment of accuracy relies on the assumption that the actual distributions in the population are similar to the information available to researchers (which is not always true).

REFERENCES

- An asterisk indicates that the reference was one of the 55 studies included in the research synthesis.*
- *Adjaye-Gbewonyo, D., Bednarczyk, R. A., Davis, R. L., & Omer, S. B. (2014). Using the Bayesian Improved Surname Geocoding Method (BISG) to create a working classification of race and ethnicity in a diverse managed care population: A validation study. *Health Services Research, 49*(1), 268–283. <https://doi.org/10.1111/1475-6773.12089>
- *Aspinall, P. J. (2001). Operationalising the collection of ethnicity data in studies of the sociology of health and illness. *Sociology of Health & Illness, 23*(6), 829–862. <https://doi.org/10.1111/1467-9566.00277>
- *Aspinall, P. J. (2003). Who is Asian? A category that remains contested in population and health research. *Journal of Public Health, 25*(2), 91–97. <https://doi.org/10.1093/pubmed/fdg021>
- *Aspinall, P. J. (2009). The future of ethnicity classifications. *Journal of Ethnic and Migration Studies, 35*(9), 1417–1435. <https://doi.org/10.1080/13691830903125901>
- Battle-Baptiste, W., & Rusert, B. (2018). *WEB Du Bois's data portraits: Visualizing Black America*. Chronicle Books.
- *Bratter, J. (2007). Will “multiracial” survive to the next generation? The racial classification of children of multiracial parents. *Social Forces, 86*(2), 821–849. <https://doi.org/10.1093/sf/86.2.821>
- *Buescher, P. A., Gizlice, Z., & Jones-Vessey, K. A. (2005). Discrepancies between published data on racial classification and self-reported race: Evidence from the 2002 North Carolina Live Birth Records. *Public Health Reports, 120*(4), 393–398. <https://doi.org/10.1177/003335490512000406>
- *Campbell, M. E. (2007). Thinking outside the (black) box: Measuring Black and multi-racial identification on surveys. *Social Science Research, 36*(3), 921–944. <https://doi.org/10.1016/j.ssresearch.2006.07.001>
- *Campbell, M. E., & Rogalin, C. L. (2006). Categorical imperatives: The interaction of Latino and racial identification. *Social Science Quarterly, 87*(5), 1030–1052. <https://doi.org/10.1111/j.1540-6237.2006.00414.x>
- Carbado, D. W., & Roithmayr, D. (2014). Critical race theory meets social science. *Annual Review of Law and Social Science, 10*(1), 149–167. <https://doi.org/10.1146/annurev-law-socsci-110413-030928>
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research, 59*, 1–12. <https://doi.org/10.1016/j.ssresearch.2016.04.015>

- *Covarrubias, A., Nava, P. E., Lara, A., Burciaga, R., Vélez, V. N., & Solorzano, D. G. (2018). Critical race quantitative intersections: A testimonio analysis. *Race, Ethnicity and Education, 21*(2), 253–273. <https://doi.org/10.1080/13613324.2017.1377412>
- Covarrubias, A., & Vélez, V. (2013). Critical race quantitative intersectionality. In M. Lynn & A. D. Dixon (Eds.), *Handbook of critical race theory in education*. Routledge Handbooks Online. <https://doi.org/10.4324/9780203155721.ch20>
- *Craemer, T. (2010). Ancestral ambivalence and racial self-classification change. *Social Science Journal, 47*(2), 307–325. <https://doi.org/10.1016/j.soscij.2009.11.001>
- *Davis, S. N., Jackson, R., & Aicardi, C. (2012). “What race do you consider yourself?” Factors influencing use of color in racial self-classification. *Sociation Today, 10*(1). <http://www.ncsociology.org/sociationtoday/v101/race.htm>
- Denton, N. A., & Deane, G. D. (2010). Researching race and ethnicity: Methodological issues. In P. H. Collins & J. Solomos (Eds.), *The SAGE handbook of race and ethnic studies*. Sage. <https://doi.org/10.4135/9781446200902.n5>
- *DiPietro, S. M., & Bursik, R. J. (2012). Studies of the new immigration: The dangers of pan-ethnic classifications. *ANNALS of the American Academy of Political and Social Science, 641*(1), 247–267. <https://doi.org/10.1177/0002716211431687>
- Dixon-Román, E. J. (2017). *Inheriting possibility: Social reproduction and quantification in education*. University of Minnesota Press.
- DuBois, W. E. B. (1899). *The Philadelphia Negro: A social study* (No. 14). University of Pennsylvania Press.
- *Eisenhower, A., Suyemoto, K., Lucchese, F., & Canenguez, K. (2014). “Which box should I check?” Examining standard check box approaches to measuring race and ethnicity. *Health Services Research, 49*(3), 1034–1055. <https://doi.org/10.1111/1475-6773.12132>
- *Elliott, M. N., Fremont, A., Morrison, P. A., Pantoja, P., & Lurie, N. (2008). A new method for estimating race/ethnicity and associated disparities where administrative records lack self-reported race/ethnicity. *Health Services Research, 43*(5 Pt. 1), 1722–1736. <https://doi.org/10.1111/j.1475-6773.2008.00854.x>
- *Feliciano, C., & Rumbaut, R. G. (2018). Varieties of ethnic self-identities: Children of immigrants in middle adulthood. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 4*(5), 26–46. <https://doi.org/10.7758/rsf.2018.4.5.02>
- *Feliciano, C., & Rumbaut, R. G. (2019). The evolution of ethnic identity from adolescence to middle adulthood: The case of the immigrant second generation. *Emerging Adulthood, 7*(2), 85–96. <https://doi.org/10.1177/2167696818805342>
- Figlio, D., Karbownik, K., & Salvanes, K. (2017). The promise of administrative data in education research. *Education Finance and Policy, 12*(2), 129–136. https://doi.org/10.1162/EDFP_a_00229
- *Fiscella, K., & Fremont, A. M. (2006). Use of geocoding and surname analysis to estimate race and ethnicity. *Health Services Research, 41*(4 Pt. 1), 1482–1500. <https://doi.org/10.1111/j.1475-6773.2006.00551.x>
- Ford, K. S. (2019). Observer-identification: A potential threat to the validity of self-identified race and ethnicity. *Educational Researcher, 48*(6), 378–381. <https://doi.org/10.3102/0013189X19860803>
- *Foy, S. L., Ray, V., & Hummel, A. (2017). The shade of a criminal record: Colorism, incarceration, and external racial classification. *Socius: Sociological Research for a Dynamic World, 3*. Advance online publication. <https://doi.org/10.1177/2378023116689567>
- *Fremont, A., Weissman, J. S., Hoch, E., & Elliott, M. N. (2016). When race/ethnicity data are lacking. *Rand Health Quarterly, 6*(1). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5158280/>

- *Garcia, J. A. (2017). The Race Project: Researching race in the social sciences researchers, measures, and scope of studies. *Journal of Race, Ethnicity and Politics*, 2(2), 300–346. <https://doi.org/10.1017/rep.2017.15>
- *Garcia, J. A., Sanchez, G. R., Sanchez-Youngman, S., Vargas, E. D., & Ybarra, V. D. (2015). Race as lived experience. *Du Bois Review: Social Science Research on Race*, 12(2), 349–373. <https://doi.org/10.1017/S1742058X15000120>
- Garcia, N. M., López, N., & Vélez, V. N. (2018). QuantCrit: Rectifying quantitative methods through critical race theory. *Race Ethnicity and Education*, 21(2), 149–157. <https://doi.org/10.1080/13613324.2017.1377675>
- *Garcia, N. M., & Mayorga, O. J. (2018). The threat of unexamined secondary data: A critical race transformative convergent mixed methods. *Race Ethnicity and Education*, 21(2), 231–252. <https://doi.org/10.1080/13613324.2017.1377415>
- Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: Education, policy, “Big Data” and principles for a critical race theory of statistics. *Race, Ethnicity and Education*, 21(2), 158–179. <https://doi.org/10.1080/13613324.2017.1377417>
- Grissom, J. A., & Redding, C. (2015). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, 2(1). <https://doi.org/10.1177/2332858415622175>
- *Grundmeier, R. W., Song, L., Ramos, M. J., Fiks, A. G., Elliott, M. N., Fremont, A., Pace, W., Wasserman, R. C., & Localio, R. (2015). Imputing missing race/ethnicity in pediatric electronic health records: Reducing bias with use of U.S. Census location and surname data. *Health Services Research*, 50(4), 946–960. <https://doi.org/10.1111/1475-6773.12295>
- Haney López, I. (2005). Race on the 2010 census: Hispanics & the shrinking White majority. *Daedalus*, 134(1), 42–52. <https://doi.org/10.1162/0011526053124479>
- *Harris, D. R., & Sim, J. J. (2002). Who is multiracial? Assessing the complexity of lived race. *American Sociological Review*, 67(4), 614–627. <https://doi.org/10.2307/3088948>
- *Hitlin, S., Brown, J. S., & Elder, G. H. (2006). Racial self-categorization in adolescence: Multiracial development and social pathways. *Child Development*, 77(5), 1298–1308. <https://doi.org/10.1111/j.1467-8624.2006.00935.x>
- *Hitlin, S., Brown, J. S., & Elder, G. H. (2007). Measuring Latinos: Racial vs. ethnic classification and self-understandings. *Social Forces*, 86(2), 587–611. <https://doi.org/10.1093/sf/86.2.587>
- *Idossa, D., Duma, N., Chekhovskiy, K., Go, R., & Ailawadhi, S. (2018). Commentary: Race and ethnicity in biomedical research: Classifications, challenges, and future directions. *Ethnicity & Disease*, 28(4), 561–564. <https://doi.org/10.18865/ed.28.4.561>
- *James, A. (2001). Making sense of race and racial classification. *Race and Society*, 4(2), 235–247. [https://doi.org/10.1016/S1090-9524\(03\)00012-3](https://doi.org/10.1016/S1090-9524(03)00012-3)
- Jones, R. S. (1998). Proving Blacks inferior: The sociology of knowledge. In J. A. Ladner (Ed.), *The death of White sociology: Essays on race and culture* (pp. 114–135). Black Classic Press.
- *Kilty, K., M. (2004). What’s in a name? Racial and ethnic classifications and the meaning of Hispanic/Latino in the United States. *Ethnic Studies Review*, 27(1), 32–56. <https://doi.org/10.1525/esr.2004.27.1.32>
- King, J. E. (2016). We may well become accomplices: To rear a generation of spectators is not to educate at all. *Educational Researcher*, 45(2), 159–172. <https://doi.org/10.3102/0013189X16639046>
- Kobrin, J. L., Sathy, V., & Shaw, E. J. (2007). *A historical view of subgroup performance differences on the SAT Reasoning Test* (Research Report No. 2006-5). College Board.

- *Kressin, N. R., Bei-Hung, C., Hendricks, A., & Kazis, L. E. (2003). Agreement between administrative data and patients' self-reports of race/ethnicity. *American Journal of Public Health, 93*(10), 1734–1739. <https://doi.org/10.2105/AJPH.93.10.1734>
- Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education, 11*(1), 7–24. <https://doi.org/10.1080/095183998236863>
- Ladson-Billings, G. (2012). Through a glass darkly: The persistence of race in education research & scholarship. *Educational Researcher, 41*(4), 115–120. <https://doi.org/10.3102/0013189X12440743>
- Ladson-Billings, G., & Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record, 97*(1), 47–68.
- *Landale, N. S., & Oropesa, R. S. (2002). White, Black, or Puerto Rican? Racial self-identification among mainland and island Puerto Ricans. *Social Forces, 81*(1), 231–254. <https://doi.org/10.1353/sof.2002.0052>
- *Lee, C. (2009). “Race” and “ethnicity” in biomedical research: How do scientists construct and explain differences in health? *Social Science & Medicine, 68*(6), 1183–1190. <https://doi.org/10.1016/j.socscimed.2008.12.036>
- Liebler, C. A., Porter, S. R., Fernandez, L. E., Noon, J. M., & Ennis, S. R. (2017). America's churning races: Race and ethnicity response changes between Census 2000 and the 2010 Census. *Demography, 54*(1), 259–284. <https://doi.org/10.1007/s13524-016-0544-0>
- *Ma, I. W. Y., Khan, N. A., Kang, A., Zalunardo, N., & Palepu, A. (2007). Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. *Journal of Clinical Epidemiology, 60*(6), 572–578. <https://doi.org/10.1016/j.jclinepi.2006.11.009>
- *Maizlish, N., & Herrera, L. (2006). Race/ethnicity in medical charts and administrative databases of patients served by community health centers. *Ethnicity & Disease, 16*(2), 483–487.
- *Marquardt, K. L., & Herrera, Y. M. (2015). Ethnicity as a variable: An assessment of measures and data sets of ethnicity and related identities. *Social Science Quarterly, 96*(3), 689–716. <https://doi.org/10.1111/ssqu.12187>
- *Mateos, P. (2007). A review of name-based ethnicity classification methods and their potential in population studies. *Population, Space and Place, 13*(4), 243–263. <https://doi.org/10.1002/psp.457>
- *Mayhew, M. J., & Simonoff, J. S. (2015a). Effect coding as a mechanism for improving the accuracy of measuring students who self-identify with more than one race. *Research in Higher Education, 56*(6), 595–600. <https://doi.org/10.1007/s11162-015-9364-0>
- *Mayhew, M. J., & Simonoff, J. S. (2015b). Non-White, no more: Effect coding as an alternative to dummy coding with implications for higher education researchers. *Journal of College Student Development, 56*(2), 170–175. <https://doi.org/10.1353/csd.2015.0019>
- *Mays, V. M., Ponce, N. A., Washington, D. L., & Cochran, S. D. (2003). Classification of race and ethnicity: Implications for public health. *Annual Review of Public Health, 24*, 83–110. <https://doi.org/10.1146/annurev.publhealth.24.100901.140927>
- McGuinn, P. (2015). Schooling the state: ESEA and the evolution of the U.S. Department of Education. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 1*(3), 77–94. <https://doi.org/10.7758/rsf.2015.1.3.04>
- Mihoko Doyle, J., & Kao, G. (2007). Are racial identities of multiracials stable? Changing self-identification among single and multiple race individuals. *Social Psychology Quarterly, 70*(4), 405–423. <https://doi.org/10.1177/019027250707000409>
- Monroe, C. R. (2013). Colorizing educational research: African American life and schooling as an exemplar. *Educational Researcher, 42*(1), 9–19. <https://doi.org/10.3102/0013189X12469998>

- *Mora, G. C. (2014). Cross-field effects and ethnic classification: The institutionalization of Hispanic panethnicity, 1965 to 1990. *American Sociological Review, 79*(2), 183–210. <https://doi.org/10.1177/0003122413509813>
- *Moscou, S., Anderson, M. R., Kaplan, J. B., & Valencia, L. (2003). Validity of racial/ethnic classifications in medical records data: An exploratory study. *American Journal of Public Health, 93*(7), 1084–1086. <https://doi.org/10.2105/AJPH.93.7.1084>
- Pang, V. O., Han, P. P., & Pang, J. M. (2011). Asian American and Pacific Islander students: Equity and the achievement gap. *Educational Researcher, 40*(8), 378–389. <https://doi.org/10.3102/0013189X11424222>
- *Parker, J. D., Schenker, N., Ingram, D. D., Weed, J. A., Heck, K. E., & Madans, J. H. (2004). Bridging between two standards for collecting information on race and ethnicity: An application to Census 2000 and vital rates. *Public Health Reports, 119*(2), 192–205. <https://doi.org/10.1177/003335490411900213>
- *Prewitt, K. (2005). Racial classification in America: Where do we go from here? *Daedalus, 134*(1), 5–17. <https://doi.org/10.1162/0011526053124370>
- *Prewitt, K. (2018). The census race classification: Is it doing its job? *ANNALS of the American Academy of Political and Social Science, 677*(1), 8–24. <https://doi.org/10.1177/0002716218756629>
- Roberts, D. (2011). *Fatal invention: How science, politics, and big business re-create race in the twenty-first century*. New Press.
- Rodriguez, C. E. (2000). *Changing race: Latinos, the census, and the history of ethnicity in the United States*. New York University Press.
- *Roth, W. D. (2005). The end of the one-drop rule? Labeling of multiracial children in Black intermarriages. *Sociological Forum, 20*(1), 35–67. <https://doi.org/10.1007/s11206-005-1897-0>
- *Roth, W. D. (2010). Racial mismatch: The divergence between form and function in data for monitoring racial discrimination of Hispanics: Racial mismatch. *Social Science Quarterly, 91*(5), 1288–1311. <https://doi.org/10.1111/j.1540-6237.2010.00732.x>
- Sablan, J. R. (2019). Can you really measure that? Combining critical race theory and quantitative methods. *American Educational Research Journal, 56*(1), 178–203. <https://doi.org/10.3102/0002831218798325>
- Santelices, M. V., & Wilson, M. (2010). Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning. *Harvard Educational Review, 80*(1), 106–134. <https://doi.org/10.17763/haer.80.1.j94675w001329270>
- *Saperstein, A., Kizer, J. M., & Penner, A. M. (2016). Making the most of multiple measures: Disentangling the effects of different dimensions of race in survey research. *The American Behavioral Scientist, 60*(4), 519. <https://doi.org/10.1177/0002764215613399>
- *Schenker, N., & Parker, J. D. (2003). From single-race reporting to multiple-race reporting: Using imputation methods to bridge the transition. *Statistics in Medicine, 22*(9), 1571–1587. <https://doi.org/10.1002/sim.1512>
- *Shah, P. R., & Davis, N. R. (2017). Comparing three methods of measuring race/ethnicity. *Journal of Race, Ethnicity and Politics, 2*(1), 124–139. <https://doi.org/10.1017/rep.2016.27>
- *Smith, N., Iyer, R. L., Langer-Gould, A., Getahun, D. T., Strickland, D., Jacobsen, S. J., Chen, W., Derose, S. F., & Koebnick, C. (2010). Health plan administrative records versus birth certificate records: Quality of race and ethnicity information in children. *BMC Health Services Research, 10*, 316. <https://doi.org/10.1186/1472-6963-10-316>
- Stoskopf, A. (1999). The forgotten history of eugenics. *Rethinking Schools, 13*(3), 12–13.
- Tatum, B. D. (2017). *Why are all the Black kids sitting together in the cafeteria? And other conversations about race*. Basic Books.

- *Thompson, C. A., Boothroyd, D. B., Hastings, K. G., Cullen, M. R., Palaniappan, L. P., & Rehkopf, D. H. (2018). A multiple-imputation “forward bridging” approach to address changes in the classification of Asian race/ethnicity on the US death certificate. *American Journal of Epidemiology*, 187(2), 347–357. <https://doi.org/10.1093/aje/kwx215>
- *Tzioumis, K. (2018). Demographic aspects of first names. *Scientific Data*, 5, Article 180025. <https://doi.org/10.1038/sdata.2018.25>
- U.S. Census Bureau, Commerce Department. (2018). *2018 End-to-end census test: Race and Hispanic origin questions* (OMB Control Number: 0607-0999).
- *Vargas, N., & Kingsbury, J. (2016). Racial identity contestation: Mapping and measuring racial boundaries. *Sociology Compass*, 10(8), 718–729. <https://doi.org/10.1111/soc4.12395>
- *Voicu, I. (2018). Using first name information to improve race and ethnicity classification. *Statistics and Public Policy*, 5(1), 1–13. <https://doi.org/10.1080/2330443X.2018.1427012>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$.” *The American Statistician*, 73(Suppl. 1), 1–19. <https://doi.org/10.1080/00031305.2019.1583913>
- *Williams, M., & Husk, K. (2013). Can we, should we, measure ethnicity? *International Journal of Social Research Methodology*, 16(4), 285–300. <https://doi.org/10.1080/13645579.2012.682794>
- Winfield, A. G. (2007). *Eugenics and education in America: Institutionalized racism and the implications of history, ideology, and memory* (Vol. 18). Peter Lang.
- Zuberi, T. (2001). *Thicker than blood: How racial statistics lie*. University of Minnesota Press.
- Zuberi, T., & Bonilla-Silva, E. (2008). *White logic, White methods: Racism and methodology*. Rowman & Littlefield.