DRAFT
Please do not quote or cite without consulting the authors first.

Select All That Apply?
The Pitfalls of Various Racial Classification Schemes in Higher Education Research

Karen Kurotsuchi Inkelas
University of Maryland

Matthew Soldner
University of Maryland

Katalin Szelényi
University of Massachusetts, Boston

Paper presented at the Annual Meeting of the American Educational Research Association
April 2009, San Diego, CA

This research was supported by grants from the National Science Foundation (#01-5-24618), The Association of College and University Housing Officers International, ACPA, and NASPA

Contact Information:

Karen Kurotsuchi Inkelas, PhD
Associate Professor
3214 Benjamin Building
University of Maryland
College Park, MD 20742
Email: kinkelas@umd.edu

Select All That Apply?
The Pitfalls of Various Racial Classification Schemes in Higher Education Research

Abstract

In this study, we examined the statistical consequences of various racial/ethnic categorization strategies employed by higher education survey researchers, and found that both descriptive and multivariate analyses can be dramatically altered based on how a researcher classifies students of bi-racial backgrounds. We discuss the implications of the usage of the various classification strategies, and conclude with recommendations for future research.

Select All That Apply?
The Pitfalls of Various Racial Classification Schemes in Higher Education Research

Introduction

For the majority of his life, President Barack Obama was permitted to select only one option for racial self-identification on the United States Census, despite his bi-racial heritage. It was not until 2000 that the U.S. Census offered millions of Americans, including President Obama, the option to self-identify as belonging to more than one racial/ethnic group for the first time in its history. As a result of this change in policy, over 6.3 million respondents (almost 2.2 percent of the population) identified themselves as belonging to two racial categories (United States Census, 2001). Moreover, 4.9% of children under the age of five were reported on the 2000 Census to be from two or more races (Lopez, 2003), a segment of the population representing the traditionally-aged college-going population for as early as 2013. Additionally, the proportion of multi-racial children varies greatly by regions of the U.S., with as much as 23 to 25% of the population under 18 in the Pacific/West and Southwest regions composed of individuals from more than one race (Lopez). Thus, higher education researchers working with demographic data, particularly in these regions, must wrestle with the thorny issue of how to classify and work with mixed-race students who will no doubt represent a significant portion of college enrollments in the near future.

The change in the race/ethnicity question on the 2000 U.S. Census evoked both support and opposition. Proponents argued that the policy shift honored the multiple heritages of a growing proportion of Americans, no longer forcing them to arbitrarily choose one identity over another for the sake of a survey. Others contended the benefits gained through multiracial identification were overshadowed by their costs. Options that allowed individuals to indicate their multi-racial heritages could lead to greater racial/ethnic differentiation and thus shrink monoracial minority representation, which in turn could have implications for legislative redistricting and the enforcement of civil rights and anti-discrimination interventions (Renn & Lunceford, 2004). Furthermore, the change would complicate ordinary users' attempts to compare generations of data collected before its implementation to present-day (and future) findings (Renn & Lunceford).

In addition to the personal and policy-related implications noted above, allowing respondents the opportunity to indicate more than one racial/ethnic background on a questionnaire poses measurement dilemmas for researchers collecting race/ethnicity data for a variety of purposes—from the U.S. census to a campus-based survey. First, how should the race/ethnicity question on the questionnaire itself be structured, and, once collected, how should the resulting responses be aggregated in statistical analyses? For example, on a survey that allows respondents to identify with multiple racial categories, should a person marking both African American and White be classified in the overall category of multi-racial or should a separate biracial African American-White category be created? Current higher education scholarship suggests that researchers have adopted idiosyncratic approaches to both posing the race question on surveys and reporting the results for multi-racial individuals (Johnson et al., 1997; Lopez, 2003). Not surprisingly, authors investigating this topic have concluded that the application of different classification schemes to the same data set can produce a series of demographic portraits that appear to depict populations with drastically different racial and ethnic make-ups (Johnson et al.; Lopez) An important question that emerges, then, is whether usage of different racial/ethnic classifications with respect to persons with mixed-race backgrounds can result in significantly different empirically-derived *findings*. This study addresses this heretofore unanswered question by scrutinizing the statistical consequences of the usage of different race/ethnicity classification schemes. As a result, we interrogate the analytic implications of the various classification strategies, and conclude with a discussion of the advantages and disadvantages of each strategy.

Background and Context of the Race/Ethnicity Classification

Historical and Policy-Related Context

The official classification of race in the United States traces back to the 18th century, and, through the years, has gone through many machinations. Lee (1993) provides a comparison of U.S. Census classifications from 1890 to 1990, and showed that the 1890 Census collected data on 8 different groups: White, Black, Mulatto, Quadroon, Octoroon, Chinese, Japanese, and Indian. Over the next 50 years, the terminology shifted with each new Census: For example, the different Black/White permutations slowly disappeared, additional Asian ethnic categories appeared (e.g., Filipino, Korean), and “Indian” was changed to “American Indian.” Beginning in the 1960s, the federal government began collecting race and ethnicity data to be used in conjunction with Civil Rights Act enforcement purposes, such as the desegregation of American public schools (Renn & Lunceford, 2004). However, it was not until 1993, after pressure from multiracial constituencies, that the Office of Management and Budget (OMB) began a formal review of federal standards for the collection and reporting of race/ethnicity data. This resulted in the October 1997 OMB standards—to apply to all federally collected data—that required that questionnaires include the five major race categories (White, Black, Hispanic, American Indian/Alaskan Native, and Asian or Pacific Islander) and the option that allows respondents to be able to identify as being from more than one category. Hence, the 2000 U.S. Census, for the first time in its history, allowed individuals to check more than one option for the race question.

However, as discussed previously, the *collection* of race/ethnicity data is one issue, but the *reporting* of those data is entirely another. Here, the federal government actually offers conflicting guidance, with different agencies adopting different strategies. On one end of the spectrum, as of 2011, the U.S. Department of Education will require all Title IV institutions to report racial/ethnic data about their student populations to the Integrated Postsecondary Education Data System (IPEDS) using the following categories, which are to be considered mutually exclusive: (a) Nonresident alien, (b) ethnic Hispanics/Latinos of any race, (c) American Indian or Alaska Native, (d) Asian, (e) Black or African American, (f) Native Hawaiian or Other Pacific Islander, (g) White, (h) two or more races, and (i) race and ethnicity unknown (National Center for Education Statistics, 2008).

Meanwhile, the Office of Management and Budget (OMB), in OMB Bulletin No. 00-02 on March 9, 2000, advocated the following strategy when analyzing data with respondents who identified as having more than one racial/ethnic background: (a) include those who indicated membership in one of the five single race categories (White, American Indian/Alaskan Native, Black/African American, Native Hawaiian/Other Pacific Islander, and Hispanic/Latino) (b) include the four most prevalent double race combinations (Native American/White, Asian/White, African American/White, and Native American/African American), and (c) include all other combinations that represent more than one percent of the sample population.

Various Race/Ethnicity Classification Schemes

Outside the federal government, it is likely that numerous approaches have been adopted in wording the race/ethnicity demographic question on the multitudinous surveys used to collect data on college campuses across the United States. Subsequently, how those data are used to create racial/ethnic categories is, no doubt, as varied as the number of different types of survey questions that exist. At least two studies have examined how different racial/ethnic classification schemes could result in radically different proportions of racial/ethnic representations. Using the 2000 U.S. Census data, Lopez (2003) categorized respondents using several different classification schemes. When considering Hispanic populations in her analysis, she found that approximately 25% of individuals under age 18 would be placed in a "catch-all" mixed-race category. Disaggregating the mixed-race category for those under 18, Lopez found that the majority (16.7%) were of White and Hispanic origin, with other combinations between 0.1 and 6.2 percent. With the same sample, Lopez then recalculated what the proportions would have been using fractional assignment (i.e., if a person reported him/herself as both White and Asian, he/she would be counted as 0.5 White and 0.5 Asian). Under this model, all forms of multi-racial classifications disappeared, and the proportion of White respondents rose about 10 percent. Finally, she calculated the proportions using multiple group assignment, where respondents are counted in each applicable category for which they respond affirmatively. In addition to causing all tabulations to exceed 100% for each racial/ethnic category, this strategy resulted in an inflation of the White and Hispanic categories by almost 20 percent.

A second study examined the impact of various forms of racial representation on a questionnaire on self-identification. Johnson et al. (1997) interviewed 69 multiracial women between 18 and 44 years of age, asking them to complete three different versions of the race/ethnicity question over the course of the interview. The first version provided only five racial/ethnic designations and asked respondents to select only one option. The second version provided the five racial/ethnic designations and an "other (please specify)" option. The third version provided the five designations, the "other" option, and a "multi-racial" option. When given the opportunity on the second form, 56.5 percent of the sample—who, on the first version provided one of the five mono-racial responses—opted to select the “other” option. The same women, when provided a “multiracial” option on the third version, chose that option 78.3% percent of the time. Interestingly, when asked which version of the race/ethnicity question provided the most accurate self-assessment of their identity, 68.9% indicated that the third form was their preference. This last finding in Johnson et al. (1997) calls to mind another dimension in the already complex issue of racial/ethnic classification, namely the thought process of individuals as they determine how to respond to a race question on a survey.

Multiracial Identity Development Models

Indeed, the process of racial-ethnic identification is not fixed across an individual’s life span. Individuals of mixed-race backgrounds may in fact change in their perspectives of their identities over time. Lopez (2003) has noted that self-identification can be influenced by a number of factors, including an individual’s phenotype, surname used, and family and community socialization. A number of theories attempt to capture the process of bi- or multi-racial identity development over time. Salient to this study, many of these theories posit that much of the questioning about one’s social identity occurs in late adolescence, the same period of life of the traditionally-aged college student.

Early identity development models proposed for mixed-race individuals (e.g., Poston, 1990) followed a stage-wise progression from an initially mono-racial form to a crisis that forces the individual to rethink his/her identity and finally to a more holistic and integrated form. However, more recent work (e.g., Renn, 2000, 2004) follows a more ecological approach, in which individuals’ conceptions of their identities are interrelated to the social environments around them. Through qualitative research conducted at several postsecondary institutions, Renn (2008) identified five patterns of multiracial identity:

1. Monoracial identity: the student chooses to identify with only one of his/her racial heritages;
2. Multiple mono-racial identities: the student identifies with different mono-racial identities at different times, depending upon the context he/she is in;
3. Multi-racial identity: the student chooses an identity that does not privilege one identity over the other, but instead is its own distinct "multi-racial" category;
4. Extra-racial identity: the student chooses not to identify with any racial/ethnic heritage, usually because he/she considers doing so to be an acquiescence to the dominant majority's viewpoint; and
5. Situational identity: the student manifests a fluid identity pattern, where the core identity is stable, but different elements become more salient in some contexts than in others (pp. 16-17).

These identity patterns may have meaningful consequences for survey response. For example, a Black/White individual may choose to select only the "Black/African American" choice on the questionnaire if he/she identifies most closely with the Black community. On the other hand, a different Black/White individual may choose the "multi-racial" option on a survey if he/she identifies most strongly as bi-racial. Finally, yet another Black/White individual may choose to skip the question altogether if he/she considers him/herself extra-racial.

Renn (2008) asserted that there were three factors that may influence how mixed-race college students may identify: physical appearance, cultural knowledge, and peer culture. How a multiracial person looks (e.g., his or her skin tone, hair color and texture, eye color and shape), can have a strong influence on how a person identifies, as well as how he/she is identified by others. In addition, many mixed-race individuals begin college with a set of cultural knowledge acquired by his/her parental and community upbringing. This may include skills and beliefs, such as language facility and religious affiliation, as well as cultural aesthetics, such as preferences for certain cuisines and forms of artistic expression. Finally, Renn and Wijeyesinghe (2001) argued that one's peer culture is extremely important to multi-racial identity development. First, the mere availability of certain communities may influence multi-racial students' identification. For example, if there are no other or very few other individuals of mixed-race heritage in the individual's peer group, he/she may instead choose to affiliate with the available mono-racial peer group. Acceptance from different peer groups also matters a great deal. If a multi-racial student sought to gain acceptance from a mono-racial peer group (or a multi-racial one) but was excluded, this exclusion would surely have an impact on how he/she subsequently identified.

Consequently, it is important to consider that, no matter how a race/ethnicity question is worded on a survey or aggregated post-hoc, racial/ethnic identification remains a largely self-reported phenomenon that begins with individuals long before it is eventually classified by researchers. That being said, given the enormous policy implications of race in American higher education that continue to vex us in the 21st century, researchers must continue to grapple with how to represent race and ethnicity in their work. As previously stated, this study scrutinizes that process by interrogating different race/ethnicity classification schemes, and the consequences of their use in higher education research.

Method

This study utilized the 2007 National Study of Living-Learning Programs (NSLLP) survey data, developed to assess living-learning (L/L) programs' efficacy in promoting important collegiate learning outcomes. Living-learning programs are academic programs based in college residence halls designed to fuse undergraduates' in- and out-of-class experiences to facilitate an enhanced learning experience. This data has been primarily used to study differences in the contributions of key collegiate environments on student outcomes among L/L and non-L/L students. The NSLLP survey instrument, however, is ideal for this study because it provided respondents the opportunity to identify with as many races/ethnicities as desired on its racial/ethnic demographic question. Thus, is provides a model venue to simulate various race/ethnicity grouping schemes employed by survey researchers.

Sample and Data Collection

Administered via the web in Spring 2007, the NSLLP included 22,519 undergraduate student respondents at 49 predominantly white and doctoral-extensive or –intensive colleges and universities nation-wide. The sample for this study was composed of two groups: a) all or a random sample of L/L participants on the respective campus; and b) a comparison sample of students living in a residence hall who were matched to the L/L sample by gender, race/ethnicity, and academic class standing. The NSLLP survey instrument included Likert-type questions that gathered information about students' demographic characteristics, high school achievement, collegiate activities, perceptions of growth, and expectations for the future.

Variables Used in the Study

In addition to students' responses to the race/ethnicity question and our recoding of them, described below, this study used a number of single items (i.e., gender, parental education, financial aid packaging, citizenship status, family income, and SAT score) and factor-based scales in its univariate and multivariate analyses. Factor-based scales representing key living and learning environments and a student outcome that are used in our exemplar multivariate analyses were created and refined via exploratory and confirmatory factor analysis (Cronbach's alphas ranged from .676 to .878). These scales include: a) course-related faculty interaction; b) faculty mentoring; c) academically supportive residence hall climate; d) socially supportive residence hall climate; e) discussion of academic and career issues with peers; f) discussion of socio-cultural issues with peers; and g) perceptions of a smooth academic transition to college (outcome). For more information about the factor-based scales, see Appendix A. It is important to note that, although the variables selected were theoretically and statistically derived for the purposes of understanding the contributions of living-learning programs in facilitating key student outcomes, their use in this study pertains to the purpose of uncovering possible differences in results when using different race/ethnicity classification schemes.

<u>Racial/Ethnic Group Coding</u>. For this study, we simulated two popular approaches in racial/ethnic classification schemes utilized in higher education research. We also conducted a simulation using the 2000 OMB classification scheme discussed previously. The original NSLLP race/ethnicity question was worded as follows: "Are you (circle all that apply): (1) African-American/Black (not of Hispanic origin, (2) Asian or Pacific Islander (includes the Indian sub-continent), (3) American Indian or Alaskan Native, (4) Hispanic/Latino (Spanish culture or origin), (5) White/Caucasian (persons not of Hispanic origin, having origins in any of the original peoples of Europe, North Africa, or the Middle East), or (6) race/ethnicity not included above." Because of difficulty in interpretation, students who identified solely with "race/ethnicity not included" were excluded from further analysis. We termed our first classification scheme *bi-racial aggregation*, in which students who identified with only one race/ethnicity designator were assigned to that group, while students who identified with any combination of two or more designators were assigned to a group labeled "biracial." We labeled

the second classification scheme *least prevalent single-race category*, where mono-racial students were assigned to the group with which they identified, and any student who selected more than one designator was assigned to the least prevalent single-race category with which they identified. For example, a student who identified as both Asian American and Native American would be assigned to the Native American group because, of the two, that group had the fewest number of total respondents. Finally, we mirrored the 2000 OMB recommendation, which we termed *bi-racial disaggregation*, for which only those bi- or multi-racial students belonging to groups whose size constituted 1.0% of the total sample were retained for analysis, along with all mono-racial students. In the NSLLP dataset, this included students checking the following two categories: (a) Hispanic and White, (b) Asian American and White, (c) Native American and White, and (d) African American and White.

Data Analyses

We began by investigating the distributions of five socio-demographic characteristics (gender, parental education, financial aid, citizenship status, and family income) across the three different racial/ethnic categorization schemes, replicating the types of descriptive analyses commonly conducted on large national datasets. Then, in order to test whether differences in racial/ethnic categorization schemes might result in altering the outcomes of multivariate statistical tests, we conducted several block-entry OLS regression analyses, using a simple college impact (e.g., Astin, 1991) framework:

- Block one: pre-college characteristics (see above)
- Block two: participation in a living-learning program
- Block three: key collegiate environments, including peer discussions, faculty interaction, and students' perceptions of their residence hall climates.

The outcome examined in the regression analysis was students' perceptions of a smooth academic transition to college. Again, although the results of these analyses may be generally informative about L/L programs and their relationship to student outcomes, the specific purpose of this study was to explore differences in these analyses related to racial/ethnic categorization strategies.

Results

Descriptive Analyses

Three series of crosstabulations were generated using SPSS 16, each placing one of the three classification scheme's race/ethnicity designators on the vertical axes and placing the levels of one of five demographic variables, including gender, parental education level, financial aid packaging, citizenship status, and family income on the horizontal axes. For each demographic variable, two sets of proportions are compared: (a) those of mono-racial/mono-ethnic students, who are treated differently in the biracial disaggregation condition (which assign mono-racial/mono-ethnic students to the same group selected by students) than those in the least-prevalent single-race condition which (which redistribute bi- or multi-racial/ethnic students into single mono-racial/mono-ethnic categories), and (b) those of bi- or multi-racial/ethnic students, who are treated differently in the biracial disaggregation condition (which specifically report the race/ethnicity of biracial students in the sample with a prevalent at or above 1.0%) and the biracial aggregation condition (which places all bi- or multi-racial/ethnic students in a "multiracial" category).

In Table 1, we summarize how different categorization schemes influence the reporting of gender. On the left side of the table, we compare two methods: least prevalent single-race and the biracial disaggregation. To quantify the difference for each racial group, we subtract the percentages generated by the biracial disaggregation method from those generated by the least-prevalent single race method, labeling them disparities. On the right side of the table, we conduct a different comparison: the biracial aggregation method and the biracial disaggregation method. For these comparisons, percentages for "any bi- or multi-racial/ethnic students" are contrasted against the four most frequently occurring multiracial conditions in our sample: (a) students who identified as both Hispanic and White, (b) students who identified as both Asian/Pacific Islander and White, (c) students who identified as both African American and White, and (d) students who identified as both Native American and White.

As can be seen in Table 1, there is an indication that the least-prevalent single-race method creates a differential pattern of disparity between racial groups, at least when the variable under consideration is gender. The effect is greatest for students who identified as Native American, where the male/female gender split is influenced by ± 2.9%. Just as the least-prevalent single-race method misrepresents the true demographic characteristics of a sample, so too does the biracial aggregation condition. Students who identified as Native American and White were most strongly affected by inclusion within a generic "biracial" category, with their population proportions being misestimated by ± 2.6%. Irrespective of method used, all gender × racial identifier chi-squares were statistically significant at the $p \leq .001$ level, suggesting that although the proportions analyzed varied, the omnibus statistical judgments would be identical. A review of standardized residuals for each cell across models revealed a similar conclusion: although discrepancies existed, they were minor at best (±0.3).

Table 1
Gender (Mono-racial/ethnic and bi- or multi-racial/ethnic students)

Mono-racial/ethnic group	Male Disparity	Female Disparity	Bi- or multi-racial/ethnic group	Male %	Female %
Black or African American	0.4%	-0.4%	Any bi- or multi-racial/ethnic	32.3%	67.7%
Asian or Pacific Islander	1.1%	-1.1%	Hispanic/White	34.1%	65.9%
Native American	-2.9%	2.9%	Asian/White	33.2%	66.8%
Hispanic	0.4%	-0.4%	Native American/White	29.7%	70.3%
White	0%	0%	African American/White	34.2%	65.8%

Note. Disparity is calculated by subtracting the percentages generated by the biracial disaggregation method from those generated by the least-prevalent single-race method (for students reported as mono-racial). Positive disparities represent “over-estimates” by the least-prevalent single-race method.

In Tables 2 and 3, the parental education (operationalized as the higher of either a student’s mother’s or father’s reported level of education) of students from different racial/ethnic groups is contrasted. Table 2 presents its results in the disparity format (i.e., it was created by subtracting the percent generated from the biracial disaggregation method from the percent generated by the least-prevalent single-race method), while Table 3 presents raw percentages for both the aggregated “any bi- or multi-racial/ethnic” identifier along with those for students who identified as Hispanic and White, Asian/Pacific Islander and White, Native American and White, and African American and White.

Table 2
Parental education (Mono-racial/ethnic students)

Mono-racial/ethnic group	Don’t Know Disparity	HS or Less Disparity	Some College Disparity	AA Disparity	BA Disparity	MA Disparity	PhD Disparity
Black or African American	-0.1%	-0.2%	-0.4%	-0.5%	0.9%	0.3%	0.1%
Asian or Pacific Islander	0%	-1.9%	0%	-0.1%	0.8%	0.9%	0.2%
Native American	-1.5%	-3.4%	-4.9%	-9.2%	11.3%	8.1%	-0.5%
Hispanic	-.1%	-6.4%	-1.1%	0.2%	2.5%	3.3%	1.4%
White	0%	0%	0%	0%	0%	0%	0%

Table 3
Parental education (Bi- or multi-racial/ethnic students)

Bi- or multi-racial/ethnic group	Don’t know	HS or less	Some college	AAs	BA	MA	PhD
Any bi- or multi-racial/ethnic	1.1%	7.6%	13.9%	7.2%	31.6%	25.2%	13.4%
Hispanic/White	0.3%	7.3%	14.6%	8.3%	32.4%	25.1%	12.1%
Asian/White	2.0%	2.0%	10.9%	4.4%	32.0%	28.6%	20.1%
Native American/White	0%	11.4%	9.0%	9.0%	32.3%	24.9%	13.4%

African American/White	0.9%	9.8%	20.5%	5.4%	36.6%	19.6%	7.1%

As can be seen in Table 2, the least-prevalent racial category method again has relatively severe consequences for racial/ethnic groups that exhibited a high proportion of bi- or multi-racial/ethnic students relative to their non-dominant, mono-racial peers. As can be seen in Table 2, the assignment of bi- or multi-racial Native American students to a monolithic Native American category resulted in discrepancies ranging from .5% to 11.3%. Students who identified (or were identified, in the least-prevalent single-race scheme) as Hispanic were also affected, with discrepancies ranging from .1% to 6.4%. In Table 3, the effects of biracial aggregation or disaggregation are presented. There, the most notable discrepancies are the tendency toward underestimating the parental education of students who identified as Asian American and White, and the lower percentage of students who identify as African American and White who come from households headed by at least one parent with a Ph.D.

Table 4 summarizes differences in how the three racial/ethnic categorization methods studied here influenced the presentation of results related to financial aid packaging. While the least-prevalent single-race method produced relatively minor discrepancies vis-à-vis students' use of student loans, it dramatically over-estimated the percent of Native American students who reported receiving merit-based aid (almost 15%). Disaggregating bi- or multi-racial/ethnic students also revealed more nuance in students' financial aid packaging than could be seen when students were aggregated into a more generic category. For example, although bi- or multi-racial/ethnic students *overall* reported using student loans at a rate of exactly 50%, disaggregation revealed that only 41.7% of students who identified as Asian American and White took student loans while 66.4% of students who identified as African American and White did so.

Table 4
Financial aid packaging: Loans vs. Merit
(Mono-racial/ethnic and bi- or multi-racial/ethnic students)

Mono-racial/ethnic group	Loan Disparity	Merit Disparity	Bi- or multi-racial/ethnic group	Loan %	Merit %
Black or African American	-0.5%	1.0%	Any bi- or multi-racial/ethnic	50.0%	47.7%
Asian or Pacific Islander	0.8%	1.6%	Hispanic/White	48.9%	48.9%
Native American	1.3%	14.6%	Asian/White	41.7%	44.5%
Hispanic	-0.7%	2.3%	Native American/White	45.7%	56.3%
White	0%	0%	African American/White	66.4%	50.0%

Note. Disparity is calculated by subtracting the percentages generated by the biracial disaggregation condition from those generated by the least-prevalent single-race condition (for students reported as mono-racial). Positive disparities represent “over-estimates” by the least-prevalent single-race method.

Table 5 presents results related to students’ citizenship status, operationalized as whether they were born in the United States. With the exception of African American students, students from all other racial/ethnic groups are depicted as being more likely to have been born in the United States when researchers use the least-prevalent single-race method than the biracial disaggregation method. Asian or Pacific Islander students are most strongly affected, at ±5.1%, followed by Hispanic and Native American students. The biracial disaggregation method also appears to more faithfully reproduce bi- or multi-racial/ethnic students’ citizenship status: discrepancies range from 2.1% (Hispanic/White students) to 4.9% (Native American/White students).

Table 5
Citizenship status
(Mono-racial/ethnic and bi- or multi-racial/ethnic students)

Mono-racial/ethnic group	US born Disparity	Not US born Disparity	Bi- or multi-racial/ethnic group	US born %	Not US born %
Black or African American	-0.4%	0.4%	Any bi- or multi-racial/ethnic	92.6%	7.4%
Asian or Pacific Islander	5.1%	-5.1%	Hispanic/White	94.9%	5.1%
Native American	2.6%	-2.6%	Asian/White	89.9%	10.1%
Hispanic	3.9%	-3.9%	Native American/White	97.5%	2.5%
White	0%	0%	African American/White	91.2%	8.8%

Note. Disparity is calculated by subtracting the percentages generated by the biracial disaggregation condition from those generated by the least-prevalent single-race condition (for students reported as mono-racial). Positive disparities represent “over-estimates” by the least-prevalent single-race method.

The final demographic variable considered here were students' report of their family incomes, presented in Table 6 (which details a comparison of the least-prevalent single-race method and the biracial aggregation method in disparity format) and Table 7 (which details a comparison of the biracial disaggregation method and the biracial aggregation method using raw percentages). While the least-prevalent single-race method underestimates the percentage of non-White students in the lowest income bracket irrespective of race, the misrepresentation of Native American students (-19.2%) and Hispanic students (-7.6%) found in Table 6 is most severe. The biracial aggregation method, used in the construction of Table 7, also produces estimates of family income that are markedly different than the biracial disaggregation method. For example, although 37.1% of *all* biracial students come from families that make less than $75,000 per year, only 25.0% of students who identify as Asian American and White, while 53.6% of students who identify as African American and White, do so.

Table 6
Family income (Mono-racial/ethnic students)

Mono-racial/ethnic group	Less than $75,000 Disparity	$75,000 to $99,000 Disparity	$100,000 to $125,000 Disparity	$125,000 to $175,000 Disparity	$175,000 or more Disparity
Black or African American	-2.1%	0.5%	0.3%	0.2%	0.1%
Asian or Pacific Islander	-3.5%	0.4%	0.7%	1.1%	1.2%
Native American	-19.2%	-1.3%	6.3%	9.3%	5.1%
Hispanic	-7.6%	1.2%	1.5%	2.4%	2.4%
White	0%	0%	0%	0%	0%

Table 7
Family income (Bi- or multi-racial/ethnic students)

Bi- or multi-racial/ethnic group	Less than $75,000	$75,000 to $99,000	$100,000 to $125,000	$125,000 to $175,000	$175,000 or more
Any bi- or multi-racial/ethnic	37.1%	17.7%	15.9%	15.5%	13.7%
Hispanic/White	31.2%	17.3%	16.9%	18.6%	15.9%
Asian/White	25.0%	18.8%	19.2%	18.1%	18.8%
Native American/White	42.7%	15.1%	15.1%	15.6%	11.5%
African American/White	53.6%	17.3%	14.5%	9.1%	5.5%

Multivariate Analyses

As noted, a multiple regression analysis was run to compare the consequences of grouping students by race/ethnicity based upon each of the three methods identified above. The dependent variable represented students' perceptions of a smooth academic transition to college. Before reporting the results, a number of assumptions underlying multiple regression were evaluated, including normality, linearity, homoscedasticity, and the absence of excessive multicollinearity. Readers are asked to use caution when interpreting findings about students who identified as both African American and White: due to list-wise deletion, the *n* underlying the regression analyses of which these students were a part hovered near (or slightly below) the conventionally held minimum threshold for analysis of 10 observations per independent variable. For this same reason, all Native American students were excluded from multivariate analysis.

In the tables below, highly edited regression tables are presented, listing only the beta weights for each variable contained in the regression equation and its estimated statistical significance. Importantly, complete tables (and other data relevant to OLS regression, including R^2 estimates and F-values) have not been reproduced because of the focus of this investigation is not on the three dependent variables selected for analysis, but instead upon differences in independent variables' parameter estimates.

Tables 8, 9, 10, and 11 summarize our results related to students' academic transition to college. Results in Table 8 pertain to students who identified as African American or African American and White via the biracial disaggregation method, and who are identified as African American by the least-prevalent racial-aggregation method. Tables 9 and 10 summarize data related to three similarly configured groups of Asian American and Hispanic students. It should be noted that results are presented in this manner so that readers can consider the consequences of the least-prevalent racial-aggregation method: what happens analytically when one adds students who, for example, identify as *both* Asian American and White to a single, monolithic Asian American group? Table 11 compares regression results of all biracial students with their disaggregated peers.

By comparing Tables 8 through 11 it is possible to infer the extent to which the three categorization methods studied here might influence readers' statistical (or practical) interpretation of multivariate analyses. Results for students who identified (or who were assigned) with a single racial group—those presented in Tables 8, 9, and 10—suggested that the least-prevalent single-race method and the biracial disaggregation method produced similar, although not identical, beta-weights and estimated statistical significance. For example, using the least-prevalent single-race method, the strongest predictor of African American students' smooth academic transition (at least in our model, and for our students), was the extent to students engaged in course-related faculty interaction (β=.150). When one uses the biracial disaggregation condition, however, the strongest predictor is African American students' perception that their residence hall has a climate that supports academic achievement (β=.155), which is then *followed* in relative importance by course-related faculty interaction (β=.120). Similarly, regression analysis based upon the least-prevalent single-race method would suggest that parental education is unrelated to Asian American students' smooth academic transition to college ($p \leq .05$), while analysis based on the biracial disaggregation method suggests that it is not ($p \geq .05$).

Comparisons involving biracial students in Tables 8, 9, and 10 are somewhat more complicated. It is important to note that the number of observations in each sample is much smaller than those used to compute the single-race regressions. While this does not systematically influence beta weights, it does make it less likely that individual independent variables will be found to be statistically significant in the regression model. Similarly, direct comparison of beta weights across equations is not advisable, a single variable's relative importance within one equation still be compared to that same variable's importance relative to its peers in another. So, for example, the strongest predictor of a smooth academic transition to college for students who identify as both Hispanic and White was the extent to which they perceived their residence hall was socially supportive, a variable that was rated third in importance for students identified as Hispanic by the least-prevalent single-race method and was insignificant in the biracial disaggregation method.

Table 8
Smooth academic transition to college (African Americans)

	African American (OMB)		African American (Least prevalent)		African American/White (Biracial disaggreg)	
Independent Variable	β	Sig	β	Sig	β	Sig
Course-related faculty interaction	.120	*	.150	**	.371	*
Faculty mentoring	.040		.031		.011	
Res hall academically supportive	.155	**	.144	**	.066	
Res hall socially supportive	.079		.075		.049	
Parental educational attainment	-.023		-.037		-.123	
Parental Income	.017		.034		.171	
Academic peer interaction	.068		.085		.189	
Social peer interaction	-.015		-.030		-.157	
Male	.012		-.002		-.170	
L/L participant	.072	*	.070	*	.131	
SAT score	-.001		.013		.065	
U.S. citizenship	-.047		-.026		.045	

Table 9
Smooth academic transition to college (Asian or Pacific Islanders)

	Asian or Pacific Islander (OMB)		Asian or Pacific Islander (Least prevalent)		Asian or Pacific Islander/ White (Biracial disaggreg)	
Independent Variable	β	Sig	β	Sig	β	Sig
Course-related faculty interaction	.099	*	.139	***	.293	**
Faculty mentoring	.016		.013		.032	
Res hall academically supportive	.138	**	.138	***	.177	
Res hall socially supportive	.092	*	.088	*	.069	
Parental educational attainment	-.045		-.040		-.006	
Parental Income	.085	*	.062		-.060	
Academic peer interaction	.088	*	.109	**	.179	*
Social peer interaction	.047		.027		-.093	
Male	-.005		.014		.112	
L/L participant	.053		.032		-.065	
SAT score	-.018		-.013		-.005	
U.S. citizenship	-.058		-.041		.010	

Table 10
Smooth academic transition to college (Hispanic)

	Hispanic (OMB)		Hispanic (Least prevalent)		Hispanic/White (Biracial disaggreg)	
Independent Variable	β	Sig	β	Sig	β	Sig
Course-related faculty interaction	.214	***	.197	***	.219	*
Faculty mentoring	.045		.023		.012	
Res hall academically supportive	.135	*	.118	*	.066	
Res hall socially supportive	.068		.138	**	.244	***
Parental educational attainment	-.011		-.030		-.023	
Parental Income	-.016		.005		.101	
Academic peer interaction	.203	***	.148	***	.024	
Social peer interaction	-.106		-.072		.015	
Male	.023		.028		-.030	
L/L participant	-.002		.011		.039	
SAT score	.105	*	.069		.031	
U.S. citizenship	-.009		.008		-.008	

Table 11 contrasts the experience of *all* bi- or multi-racial/ethnic students (the biracial aggregation condition) with three biracial groups, students who identify as African American and White, Asian American and White, and Hispanic and White. Again, while direct comparisons of beta weights is *not* advised, the relative importance of a variable within the context of a larger system of variables can be considered across groups. For example, the biracial aggregation method would suggest that, for biracial students, the most important contributor to students' smooth transition to college is their course-related faculty interaction. Looking across models, the same appears to be true for students who identify as African American and White and Asian and White. Students who identify as Hispanic and White, however, evidence a slightly different pattern: students' residence hall social climate (β=.244) is a stronger predictor relative to course-related faculty interaction (β=.219).

Table 11
Smooth academic transition to college (Biracial students)

	All Biracial (Biracial aggreg)		African American/ White (OMB)		Asian Pacific American/ White (OMB)		Hispanic/ White (OMB)	
Independent Variable	β	Sig	β	Sig	β	Sig	β	Sig
Course-related faculty interaction	.241	***	.371	*	.293	**	.219	*
Faculty mentoring	-.005		.011		.032		.012	
Res hall academically supportive	.124	*	.066		.177		.066	
Res hall socially supportive	.164	***	.049		.069		.244	***
Parental educational attainment	-.054		-.123		-.006		-.023	
Parental Income	.052		.171		-.060		.101	
Academic peer interaction	.088		.189		.179	*	.024	
Social peer interaction	-.053		-.157		-.093		.015	
Male	.019		-.170		.112		-.030	
L/L participant	.006		.131		-.065		.039	
SAT score	.020		.065		-.005		.031	
U.S. citizenship	.036		.045		.010		-.008	

Implications for Research and Practice

Results of this study have shown that the way in which researchers choose to group students by their responses to race/ethnicity items can have profound implications for the descriptive portraits that emerge from their work as well as the multivariate analyses with which theory may be evaluated and built. Unfortunately, there is no single solution to this empirical dilemma. Indeed, each approach has its strengths and its limitations.

Clearly, the least-prevalent single-race method is beneficial for researchers seeking to maximize their statistical power through increasing the proportion of underrepresented minorities in their samples. However, this study has shown that this strategy tends to misrepresent, to varying degrees, the actual characteristics of students within each racial category. Most importantly, this misrepresentation might result in researchers' failure to recognize significant inequalities in college student populations. Treating a respondent who has marked the racial categories of Native American and White as mono-racial Native American, for example, might result in "diluting" the extent of educational disadvantage experienced by students whose mono-racial identification as Native American is driven solely by their personal choice (rather than the choice of the researcher). Undoubtedly, limiting mono-racial categories in statistical analyses strictly to those respondents who themselves identify as mono-racial results in lowering the numerical representation of minority racial groups. This, in turn, holds implications for the enactment and distribution of civil rights policies. However, from a research perspective, recognizing the true extent of inequality among underrepresented minority groups is only possible if those minority groups are not grouped together with bi- or multi-racial individuals. In addition, the arbitrary assignment of a student to a racial/ethnic category that he or she did not wholly designate seems, to us, presumptuous.

An alternative approach is the biracial aggregation method. The benefit of such an approach is that, in contrast to the least-prevalent single-race method, students who identify as bi- or multi-racial/ethnic are no longer grouped, perhaps erroneously, with single-race peers. Presumably, this improves the accuracy of estimates derived from both samples. However, aggregating all bi- or multi-racial/ethnic students into one monolithic category glosses over genuine differences between students of differing multi-racial heritages. As both univariate and multivariate analyses revealed, this technique masks real differences between students, hampering our ability to understand the complexities of the multi-racial student experience. After all, the backgrounds, educational experiences, and outcomes of Asian American/White and African American/Native American students, for example, might vary so widely that the commonality in their multi-racial heritage falls short of justifying their representation within a single "multi-racial" category. (For now, we choose to accede to practices that combine Asian American or Latino students into monolithic groups of their own, although we understand the same critique might be directed toward this form of aggregation. Migration histories and socioeconomic circumstances likely result in vastly different experiences for Latinos who identify with Cuba as opposed to Honduras, or for Asian Americans who would identify with Japan as opposed to Laos.)

The final method, the disaggregated biracial approach, has as its primary benefit the most faithful representation of students' preferred racial/ethnic identification. Offering the finest level of granularity available to the researcher, this method allows for the greatest clarity in data analysis and accuracy in the inferences drawn from statistical tests. With granularity, however, comes a not insubstantial drawback: a loss of statistical power. This may complicate, or render impossible, some types of multivariate analyses, resulting in researchers' inability to draw meaningful conclusions about the experiences and outcomes of specific bi- or multi-racial student populations. This problem is especially prominent in small datasets, where even the smaller mono-racial samples—oftentimes, Native Americans—might fail to achieve statistical power. Further disaggregating these already small samples raises considerable problems for data analysis.

The way in which bi- and multi-racial data are collected and used on college campuses also has implications for institutional practice. Choices to use certain classification strategies over others may result in the over-inflation or under-representation of racial/ethnic backgrounds represented in the student body, and thus allocations for programs and services based on the racial/ethnic proportions of the student enrollments may be misaligned. In addition, certain classification schemes may misrepresent students' backgrounds and experiences in ways that affect the effective targeting of student services. For example, re-classifying mixed-race students with a partial Native American heritage with students who list their only racial/ethnic identity as

Native American tends to increase the proportion of "Native American" students with college-educated parents. Thus, practitioners using this information may mistakenly believe that Native American students on the campus are not likely to be first-generation college students and thus may not structure their programs and services accordingly. Similarly, aggregating all bi- or multi-racial students into one "multi-racial" category may mask differences in student needs. For example, while 50 percent of bi-racial students overall in this study took out personal loans to pay for college, Asian American/White students were significantly less likely to be on student loans, and African American/White students were significantly more likely. It may be that African American/White students are taking out loans at higher rates because they are unaware of other financial aid opportunities. In this case, targeting programs and services to this population might increase their ability to access other funding opportunities. However, if the only information available was the aggregated "50 percent" statistic, practitioners would not know of the discrepancy among the various bi- or multi-racial student populations and could not direct resources and information effectively.

Ultimately, a researcher's choice likely depends upon the relative importance of the following factors: beliefs about the preservation of respondent choice and the capacity to conduct an analysis (due to sufficient *n*) versus the accuracy of those analyses. In light of this study's findings, this choice should not be made lightly, especially given the complex process of racial and ethnic identity development in which both mono-racial and multi-racial students engage (Helms, 1995; Renn, 2008). Although it is not within the scope of this paper to dictate the direction of other research projects, we recommend that respondents be given the option to identify with more than one racial/ethnic category, and that, to the extent possible, researchers honor students' choices by selecting aggregation methods (and subsequent analyses) that are both faithful to respondents and that elicit statistically meaningful results.

Directions for Future Research

We conclude this study with two recommendations for future research. Extending the work of Johnson et al. (1997), it would be interesting to transfer the decision-making process of how to categorize mixed-race individuals from researchers to the hands of the students themselves. Future inquiry might investigate if there are prevailing classification schemes that are more popularly endorsed by mixed-race individuals than others. After all, the constituency affected most significantly by the various classification schemes, one could argue, should have a stake in how they are ultimately treated. Similarly, following Renn's (2000, 2004, 2008) work on multi-racial identity development, it is possible that mixed-race students' perceptions of their racial/ethnic identities change over time and can be influenced by their college environments. Thus, future scholarship on multi-racial students might investigate if students' depictions of their identities do, indeed, change, and the environmental factors associated with such changes. Given that demographic forecasts predict that some postsecondary institutions could be welcoming student enrollments that could be as great as one-quarter multi-racial (Lopez, 2003), the more that is learned about this growing population can be of great benefit to researchers and practitioners alike.

References

Helms, J. E. (1995). An update of Helms's White and People of Color Racial Identity Development Models." In Ponterotto, J.G., Casas, J. M., Suzuki, L.A., and Alexander, C. M. (Eds.). *Handbook of Multicultural Counseling*. Thousand Oaks, CA: Sage.

Johnson, T. P., Jobe, J. B., O'Rourke, D., Sudman, S., Warnecke, R., Chavez, N., Chapa-Resendez, G., & Golden, P. (1997, December). Dimensions of self-identification among multiracial and multiethnic respondents in survey interviews. *Evaluation Review, 21*(6), 671-87.

Lee, S. M. (1993) Racial classifications in the U.S. Census: 1890-1990. *Ethnic and Racial Studies, 16*(1), 75-94.

Lopez, A. M. (2003, August-September). Mixed-race school-age children: A summary of Census 2000 data. *Educational Researcher*, 25-37.

National Center for Education Statistics. (n.d.) Resources for implementing changes to race/ethnicity reporting in IPEDS. [Electronic version]. Retrieved March 17, 2009, from http://nces.ed.gov/IPEDS/reic/resource.asp)

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. [Electronic version]. Retrieved March 17, 2009, from http://georgewbush-whitehouse.archives.gov/omb/fedreg/1997standards.html

Office of Management and Budget (2000). Guidance on aggregation and allocation of data on race for use in Civil Rights monitoring and enforcement, OMB Bulletin No. 00-02. [Electronic version]. Retrieved March 17, 2009, from http://georgewbush-whitehouse.archives.gov/omb/bulletins/b00-02.html

Poston, W. S. C. (1990). The biracial identity development model: A needed addition. *Journal of Counseling and Development, 69*, 152-55.

Renn, K. A. (2000). Patterns of situational identity among biracial and multiracial college students. *Review of Higher Education, 23*, 399-420.

Renn, K.A. (2004). *Mixed race students in college: The ecology of race, identity, and community*. Albany, NY: SUNY Press.

Renn, K. A. (2008, Fall). Research on biracial and multiracial identity development: Overview and synthesis. *New Directions for Student Services, 123*. Hoboken, NJ: Wiley Interscience.

Renn, K. A. & Lunceford, C. J. (2004). Because the numbers matter: Transforming postsecondary education data on student race and ethnicity to meet the challenges of a changing nation. *Educational Policy, 18*(5), 752-783.

Wijeyesinghe, C. L. (2001) Racial identity in multiracial people: An alternative paradigm. In Wijeyesinghe, C. L., & Jackson, B. W. (Eds.). *New perspectives on racial identity development: A theoretical and practical anthology*. New York: New York University Press.

Appendix A
Factor loadings and reliability measures

Factor name Individual item stem	Loading	Cronbach's α
Academic and vocational conversations with peers		.806
Shared concerns about classes and assignments	.816	
Discussed something learned in class	.787	
Talked about current news events	.692	
Social and cultural conversations with peers		.885
Discussions with students whose political opinions very different	.818	
Held discussions with those with different religious beliefs	.811	
Discussed social issues such as peace, human rights, justice	.791	
Discussed views about multiculturalism and diversity	.775	
Discussions with students whose personal values different	.696	
Course-related faculty interaction		.772
Visited informally with instructor before/after class	.798	
Made appt to meet instructor in his/her office	.706	
Asked instructor for info related to course	.685	
Non-course-related faculty mentorship		.738
Discussed personal problems or concerns with instructor	.810	
Discussed career plans & ambitions with instructor	.702	
Visited informally with instructor on social occasion	.587	
Supportive academic climate in residence hall		.800
Environment supports academic achievement	.837	
Most students study a lot	.705	
It's easy to form study groups	.672	
Staff helps with academics	.624	

Supportive social climate in residence hall		.878
Help and support one another	.783	
Appreciate different religions	.739	
Intellectually stimulating environment	.731	
Appreciate different races/ethnicities	.711	
Would recommend this residence hall	.709	
Different students interact with each other	.689	
Peer academic support	.648	
Perceptions of a smooth academic transition to college		.760
Ease with communicating with instructors outside of class	.804	
Ease with academic/personal help when needed	.706	
Ease with forming study groups	.542	