One Organization's Bridging Strategy for Racial and Ethnic Classification

By

C. Anthony Broh
Director of Research
Consortium on Financing Higher Education

Stephen D. Minicucci
Senior Research Associate
Consortium on Financing Higher Education

Presented at the 2003 Annual Forum
of the
Association for Institutional Research

May 21, 2003
Tampa, Florida

Annual Forum of the Association for Institutional Research
Tampa, Florida, May 21, 2003

One Organization's Bridging Strategy for Racial and Ethnic Classification

Abstract

This paper describes work in progress for creating a racial and ethnic classification schema that meets several research objectives that are relevant to the higher education community. The research staff of the Consortium on Financing Higher Education (COFHE)[1] has reviewed several prominent strategies for handling racial and ethnic data and has settled on a coding hierarchy for multiple response categories in several recent surveys. The classification system allows comparison with other data sets and provides a strategy for bridging data between the 1977 and 1997 government guidelines.

[1] The Consortium on Financing Higher Education (COFHE) is an institutionally supported organization of thirty-one private colleges and universities. COFHE has its origins in the Sloan Study Consortium which was formed in 1971 and funded by the Alfred P. Sloan Foundation to examine how selective, private colleges and universities would be financed. COFHE member institutions include Amherst College, Barnard College, Brown University, Bryn Mawr College, Carleton College, Columbia University, Cornell University, Dartmouth College, Duke University, Georgetown University, Harvard University, The Johns Hopkins University, Massachusetts Institute of Technology, Mount Holyoke College, Northwestern University, Oberlin College, Pomona College, Princeton University, Rice University, Smith College, Stanford University, Swarthmore College, Trinity College, The University of Chicago, University of Pennsylvania, The University of Rochester, Washington University in St. Louis, Wellesley College, Wesleyan University, Williams College, Yale University.

Introduction

Americans have been counting people by race and ethnicity since the U.S. Constitution specified separate tallies for "free persons" and slaves, and the exclusion of "Indians, not taxed." In practice, the antebellum Census Bureau went further than this, tracking free whites and "colored persons" (or "other free persons") separately. In the second half of the nineteenth century, the Census began to track those of mixed black parentage and also began to count Indians and newcomer Asian groups such as Chinese and Japanese.[2] Mainly, this is not a happy history. The categories reflect the social and political realities of white dominance. The labels helped to enforce social status as well as impose a lower class of citizenship, especially on African Americans, Asians and American Indians. In the words of one demographer,

> The fundamental questions were whether the differences between blacks and whites, and between immigrants and natives, were more or less permanent and therefore significant, or were transitory and thus to be discounted in a longer perspective.[3]

In the era of the Civil Rights movement, racial and ethnic labels began to be used for constructive purposes. Legally, these categories defined "suspect classifications" protected by the Fourteenth Amendment. Socially and politically, labels which had once been used to isolate and stigmatize were now embraced as mechanisms of accountability. We began to count by race and ethnicity to monitor our collective progress toward an integrated and equal society. The dangers of counting people by race are not forgotten, however, and the question of how to count remains a charged political and social issue.

The current system for counting individuals by race and ethnicity has its origins in the landmark decisions of the United States Supreme Court that tied civil rights to various racial and ethnic groupings and Federal policies that identified financial resources for various racial and ethnic groupings. In 1977, the Office of Management and Budget (OMB) issued guidelines to Federal agencies standardizing the racial and ethnic classification. OMB Statistical Directive 15, as it was known, provided for four races (white, black, Asian American, and American Indian) and one ethnic group (Hispanic) and established formats for collecting and reporting the data. With only slight modifications in nomenclature and definitions, the formula persisted for two decades.

OMB 15 guided data collection both within and outside the government. Conceptualized as a personal identification, the standards allowed individuals to select the one, and only one, racial/ethnic category from the list of possible alternatives.[4] Changing political and social sensibilities in the 1990s led the Census Bureau to allow respondents to select

---

[2] For a complete list of racial categories used by the U.S. Census, see Melissa Nobles, *Shades of Citizenship* (Stanford University Press, 2000) 28, 44.
[3] William Peterson, *Ethnicity Counts* (New Brunswick: Transaction Publishers, 1997), p. 12.
[4] Actually OMB 15 allowed for two different formats for collecting data, one of which separated race and ethnicity into two overlapping classifications. The Higher Education community, however, typically selected the format that included both race and ethnicity in one list of mutually exclusive categories.

more than one racial identity in the 2000 census.[5] This minor change in the instructions to data collection produced a nightmare for data reporting, especially for comparisons of racial and ethnic categories from one data collection period to another. How does one compare those who identify with more than one racial or ethnic category to those who never had more than one option?

The Census and OMB changes underscore additional problems for those who use racial and ethnic date to monitor higher education. One change was a division of the old category of "Asian and Pacific Islanders" into two categories: "Asians" and "Native Hawaiians and Other Pacific Islanders." The resulting problem is the creation of a new and very small category. Not only did this shift create a discontinuity in reporting categories across time, the change affected a group that is likely to represent only a small proportion of most studied populations. Very small numbers already reduced the utility of the historically significant count for American Indians and Alaska Natives. The creation of another small group challenges the utility of other aggregations. How should we report data for very small groups?

Additionally, the 1997 Standards continued what we believe to be the false distinction between race and ethnicity that was created by OMB Directive 15. Data collection formats create the distinction with the so-called "two question format" in which individuals are asked to indicate whether they are Hispanic in one question and indicate their "race" in another. This format does not conform to American's evolving understanding of racial and ethnic identity, does not conform to standard practices in higher education, and is, in any case, rendered unnecessary by the multiple response format adopted for the race question. Our third question, then, is: Should we continue to make a distinction between race and ethnicity?

The Department of Education has done little to address these three problems.[6] Several roundtable discussions and task forces, for example, failed to produce a reporting format for race and ethnicity or to form a consensus on "bridging strategies." These are necessary to answer our third question by allowing a comparison of racial and ethnic data between single identity surveys using OMB 15 and multiple identity survey items using the 1997 Guidelines. The Department of Education has, in fact, provided no guidance to institutions on how to store racial and ethic data electronically beyond the obvious statement of coding all 126[7] possible combinations.

In sum, the new OMB standards have created a series of unanswered questions for researchers who need to collect racial and ethnic data. The option for multiple responses creates new categories that must be mapped into the old. The distinction between one

[5] In addition, it modified the nomenclature and definitions of OMB 15, mandated a two question format for race and ethnicity for government agencies, and divided the Asian American classification into two groups.
[6] At this writing, staff from the National Center for Education Statistics are not even allowed to talk about racial and ethnic classification due to the highly politicized Supreme Court case about affirmative action in college admissions.
[7] The number of combinations of six racial ethnic groups (the five listed plus "other") taken one, two, three, four, five, or six at a time is 63. Each of these may be paired with Hispanic or not, for a total of 126.

ethnicity and a group of races continues a distinction that is logically inconsistent. The division of an old category into two creates distinctions that previously were not present.

While instructions for the higher education community have been conspicuously absent, an essential task for researchers is to create a classification system that addresses these three questions. Such a schema should conform to the 1997 standards and allow trend analysis for current data that are consistent with the requirements of OMB Directive 15. It should also adopt the most recent nomenclature and definitions where appropriate to higher education statistics.

The 1997 Standards and COFHE Conventions

The 1997 government standards include the following nomenclature and definitions for collection and reporting of race and ethnicity:

> **American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.
>
> **Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
>
> **Black or African American.** A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."
>
> **Hispanic or Latino.** A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin," can be used in addition to "Hispanic or Latino."
>
> **Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
>
> **White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

The list of races and ethnicity has several characteristics. First, it makes no provision for an "other" race or ethnicity. By implication, the list is therefore exhaustive.[8] Second the list also includes no category of "unknown." A common instruction in higher education is to classify individuals that do not self-identify from information contained in other administrative files or data sources about the individual. Third, the list has mutually exclusive definitions. While the collection procedures allow for people to be classified in

[8] The categories are not, however, "mutually exclusive" as will be discussed below.

more than one group, the definition of each group is distinct.[9] Fourth, the list divides the prior category of "Asian and Other Pacific Islander" into two groups: "Asian" and "Native Hawaiian or Other Pacific Islander" grouping." Finally, the word "or" is used two different ways. On the one hand, it joins together two different groups of people into a single category (e.g. American Indian *or* Alaska Native); on the other, it offers alternative labels for the same group (e.g. "Black *or* African American).

Labels are important and may differentiate social conditions or even political tendencies of racial and ethnic groups. For example, Mexican American migrant workers that organized in the California lettuce fields in the 1960s preferred the term "Chicano" to describe their ethnicity. "Hispanic" is a descriptor that Cubans or Mexican Americans or Puerto Ricans did not create to describe their background. Indeed the aggregation of Spanish speaking groups into a single ethnic category was in part a political reaction to "Chicano" and Puerto Rican nationalism in the 1960s. Until Richard Nixon used the term "Hispanic" in a 1969 speech, the "Hispanos" had always referred to Mexican nationals residing within the United States boundary at the time portions of Mexico were annexed in the mid nineteenth century. Ironically, the term actually differentiated these former Mexican nationals from other peoples living in South or Central America,[10] which are now included in a single ethnic category. The 1997 Guidelines recognized ambivalence among several of the groups about the nomenclature by adopting "Latino" as an alternative label.

Recent COFHE surveys use the 1997 nomenclature to ensure that respondents see labels that are common and widely used in government and media reporting about race and ethnicity. However, COFHE has adopted some conventions that simplify the nomenclature for reporting racial and ethnic data. (See Table 1) In general, the titles for each group truncate the name to that of the largest group in the category, or combine small groups to the OMB 15 directive. The goal is to use terminology that best captures the meaning of the data at the point of collection but also to allow interpretation of the data according to easily understood conventions across time.

In this same context, COFHE surveys typically do not have an option for "other" or "mixed" or "unknown." These categories will be discussed below; they generally result from procedures for analyzing and reporting the data after collection.

[9] Census categories in the 1880 included "octoroons" and "quadroons" and "mulatto" as nomenclature for varying amounts of "Negro" ancestry and any other racial ancestry.

[10] By defining this ethnic group by continent of origin rather than language, the census categories include people from

TABLE 1
Nomenclature for Racial and Ethnic Categories

1977 OMB 15 Directive	COFHE Reporting Nomenclature	1997 Guidelines
Native American	**American Indian**	American Indian or Alaska Native
Asian or Other Pacific Islander	**Asian**	Asian
Black or Afro-American	**Black**	Black or African American
Hispanic	**Hispanic**	Hispanic or Latino
NA	**Asian**	Native Hawaiian or Other Pacific Islander
White	**White**	White

Answering the three questions

As discussed in the introduction to this paper, the 1997 Standards and subsequent Guidelines created several problems for data collection in the higher education community: We summarize those questions as follows:

1. Should the collection format, and therefore the reporting format, make a distinction between race and ethnicity?
2. Should very small categories of race and ethnicity be included in a classification?
3. How can a classification that includes multiple responses for race and ethnicity be compared to a classification that allows only one response for race and ethnicity?

1. Should the collection format, and therefore the reporting format, make a distinction between race and ethnicity?

The category of "Hispanic" has its origins in the 1970 census when government reports began to describe people with a "Spanish surname."[11] The fact that the analysis of Hispanic people was an afterthought in the census is, perhaps, one reason that the group was not included in the racial classification devised in OMB 15. Being of Spanish origin was considered an ethnic classification while Asian, Black, White, or American Indian were considered racial classifications.

OMB 15 described two formats for collecting data about Spanish origin. The first asked individuals about their ethnicity and race with separate questions, the so-called "two question format." The second format included race and ethnicity in a single list of racial and ethnic options. While the Census Bureau and most government agencies used the former, higher education has traditionally used the latter. For example, the Integrated

[11] People listing their race, for example, as "Mexican" in prior census reports were considered "white."

Post-secondary Education Data System (IPEDS) lists "Hispanic" as one of several options in its instructions defining "race and ethnicity," the IPEDS forms designate "Hispanic" as a mutually exclusive column heading along with the other race categories.

The "Provisional Guidelines of 1997" specify that government agencies collecting primary data must use the two-question format for ethnic and racial data but do not require the higher education community to comply with this standard. The language in the guidelines encourages the Department of Education to "obtain more complete information" about Hispanic origin and, earlier in the document, declares that data "should" be collected in a two-question format. However, the guidelines again do not say "must" nor do they require the two-question format for either the enforcement of Title IV or laws against civil rights violations in higher education. Indeed, the guidelines provide alternative suggestions for tabulating data about Hispanic origin that would "minimize the burden on data collectors."[12]

There are a number of reasons to prefer a single-question format for race and ethnicity, especially with the possibility of multiple responses. First and most simply, given a multiple response format, using one question instead of two reduces the complexity of the data collection instrument but does not reduce the information contained in the data. The logic of allowing a respondent to make several choices among options is the same as allowing the respondent first to select between dichotomous categories and later to select among the remaining options. Both formats are simply multiple choices of all available options.

The Census form makes this point visually. The Hispanic selection has a similar appearance to the race selections. For example, the word "ethnicity" is not used at all in Question 7, but the word "race is used in Question 8. Note also that the appearance of the multiple response set for Question 7 is similar to the appearance for Question 8. Additionally, the choices for Question 8 do not have a one-to-one relationship to the reporting categories of the 1997 guidelines.

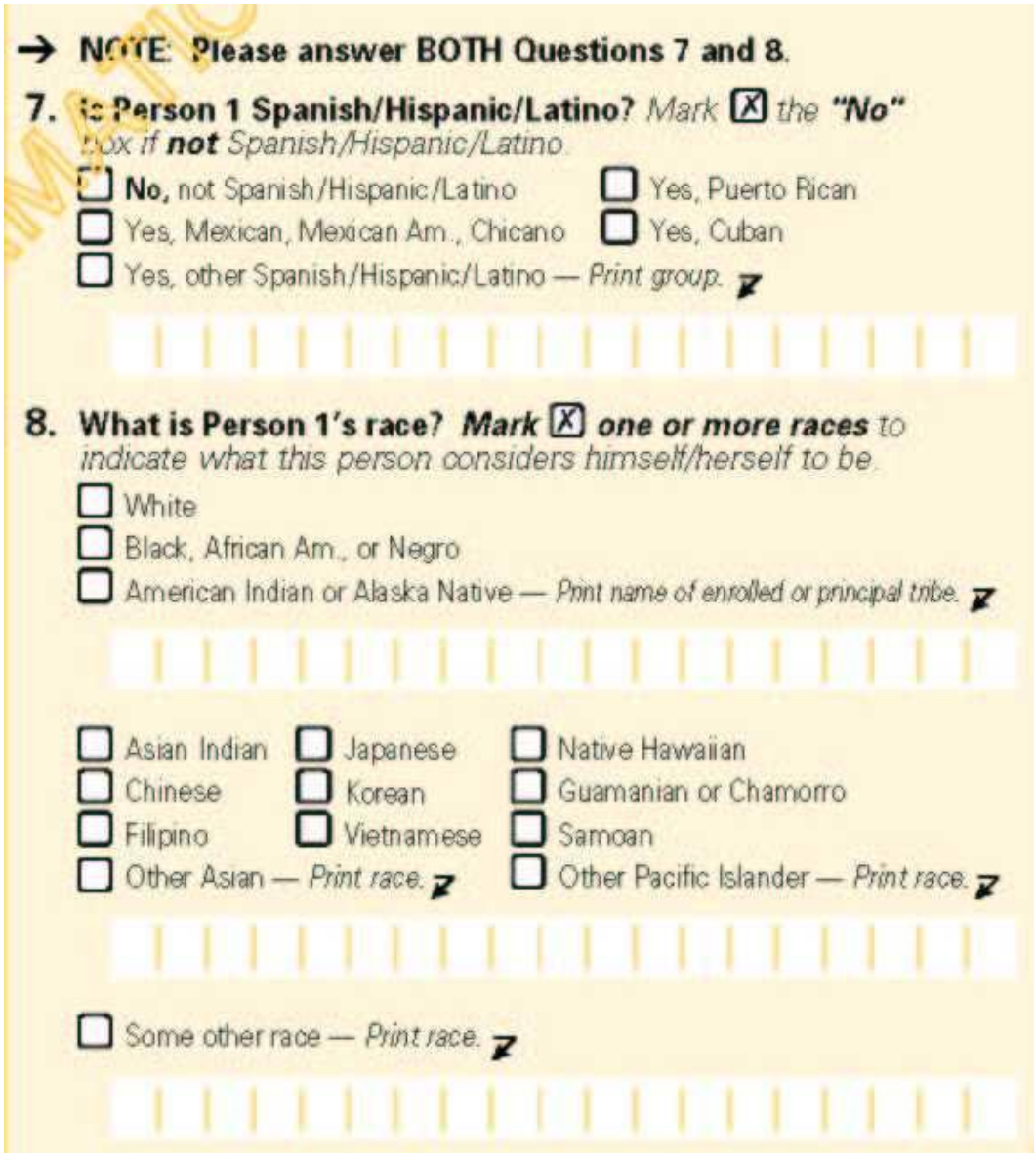
Computer technology is also a metaphor for the absence of a policy distinction between race and ethnicity Multiple choice items are typically displayed on a web survey as a "check box." The data, however, must be stored as a dichotomy – i.e. checked or not checked - to preserve the information. A "yes" or "no" might appear as a "radial button" on a web survey, but the two value response options are also a dichotomy.

More fundamentally, the single-question format, by roughly equating the concepts of race and ethnicity, comes closer to reflecting the current meaning of these terms as employed by the government, by higher education, and by ordinary individuals. The government definitions of race and ethnicity describe the classification as a psychological identification and not anthropological or biological categories. Indeed, this is the only sense in which self-identification has much meaning. Higher education admission policies typically consider "Hispanic or Latino" as one of several alternative student

[12] See C. Anthony Broh, "Summary: Provisional Guidelines on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity," *AIR Currents* 38 (Summer, 2001), p. 6.

identities. One might identify with more than one group, but psychological identification with an ethnic group is not substantively different from psychological identification with a racial group. The *Condition of Education 2002*, for example, presents "Hispanic" data as one of several "student characteristics" among other racial categories.[13]

FIGURE 1
Ethnicity and Race Question on the 2000 Census form



[13] The title of the graph is "UNDERGRADUATE DIVERSITY: Percentage of undergraduates with selected student characteristic: 1999-2000," but the graphic also includes gender and age. See, *The Condition of Education 2002,* p.99.

Finally, even a cursory look at responses to the 2000 Census suggests that Hispanic individuals often have difficulty describing themselves with the two-question format. When answering the "race" question, a substantial proportion of the Hispanic population (42 percent) chose "other." Apparently, they felt that all of the racial labels were inappropriate (a plurality do indicate "white" as their race). Since the 1997 Guidelines do not include "Other" (the categories are intended to be exhaustive), this represents a breakdown of sorts. Virtually all those who identify themselves as "other" in the 2000 Census are Hispanic (97 percent), so it would appear that most Latinos could identify themselves well with a single-question multiple-response format. [14]

For these reasons, COFHE has adopted the single-question format in its surveys that include "Hispanic or Latino" as one of several racial or ethnic choices.[15] The data are also reported in a single list of alternatives.

2. Should very small categories of race and ethnicity be included in a classification?

Prior to the 1997 Guidelines, OMB 15 combined Asian Americans with Native Hawaiians and Other Pacific Islanders; the nomenclature for this classification was "Asian and Other Pacific Islanders." The division of the OMB 15 category created parity between the treatment of the indigenous population of Hawaii and the treatment of the indigenous populations of Alaska and the other contiguous 48 states. Now Native Hawaiians would be counted in their own racial category just as American Indians or Native Alaskans had previously been counted in their own category.

Several State and Federal policies might suggest that the separate classification for Hawaiians is appropriate. For example, some schools and land use programs in Hawaii treat its indigenous population similarly to Federal programs for American Indians or Eskimos. They base program eligibility on an ancestry that dates to a time before colonization from Europe and Asia or they "grandfather" people into a program or policy based on that ancestry.[16] Additionally, the inclusion of American Indians and Eskimos as "Native Americans" with the exclusion of Native Hawaiians in a category once called "Native Americans" somehow implies that the indigenous people of this 50th state are not "Americans."

But the creation of a separate category of "Native Hawaiian and Other Pacific Islander" does little for data analysts at most institutions of higher education. First, the conceptual justification for its creation does not produce an analytic category that informs most data collection or reporting. Native Hawaiians do not, for example, behave or form opinions or engage in cultural and educational habits that are closely aligned with American Indians or Eskimos. Additionally, the creation of a separate category produces a very small racial and ethnic grouping. In a 2002 survey of the Consortium on Financing

[14] US Census Bureau, Census 2000 Summary File 3 (SF3), tables P6 and P7.
[15] Individuals may, of course, select more than one of the choices, which is the subject of "multiple selection."
[16] See Peterson, pp. 141-49.

Higher Education schools, for example, only 110 out of 41,844 racial and ethnic identifications (0.3 percent) were Native Hawaiian. In a special enumeration, the 1960 census identified only 114,405 Native Hawaiians in the state of Hawaii. Very small categories of people, when compared with larger categories of people, can result in conclusions that are not representative of the institution as a whole. While combining Native Hawaiians and other Pacific Islanders with American Indians addresses the conceptual problem, it ignores the small group problem for most colleges and universities.[17]

Additionally, government publications from the National Center for Education Statistics do not report Native Hawaiians as a separate category. For example, *The Condition of Education 2002* displays statistics from 1999-2000 about undergraduate diversity with a label of "Asian/Pacific Islander," a label that is not described in the 1997 Guidelines.[18] This nomenclature implies an inclusion of the new category, "Native Hawaiians or Other Pacific Islander" with the old "Asian or Other Pacific Islander."

Thus the COFHE Classification of Race and Ethnicity that is reported in this paper and other reports combines Native Hawaiians and Other Pacific Islanders with Asian, which is best conceptualized as the original OMB 15 category of "Asian and Pacific Islander." We adopt the current nomenclature of "Asian" as the most inclusive, albeit an imperfect,[19] descriptor, that is both short and consistent with the area of origin for the people in this category.

3. How can a classification that includes multiple responses for race and ethnicity be compared to a classification that allows only one response for race and ethnicity?

None of the previous discussion, however, allows comparison of single response items with multiple response items. It does, however, allow comparison of both the old and the new categories with recently collected data. Yet even this comparison is complicated by the choice of a bridging strategy. While the proportion of Americans who identify with two or more races is still small, 2.6 percent in the 2000 Census, it is clear that this figure is growing. In a recent survey, 1.5 percent of COFHE *parents* identified themselves as belonging to two or more racial/ethnic groups, but said that 4.2 percent of *their children* could be so identified.

Figure 2 compares the racial and ethnic identities of parents and children from the COFHE 2002 Parents survey. The discussion of parent and student race and ethnicity might be difficult to grasp.

---

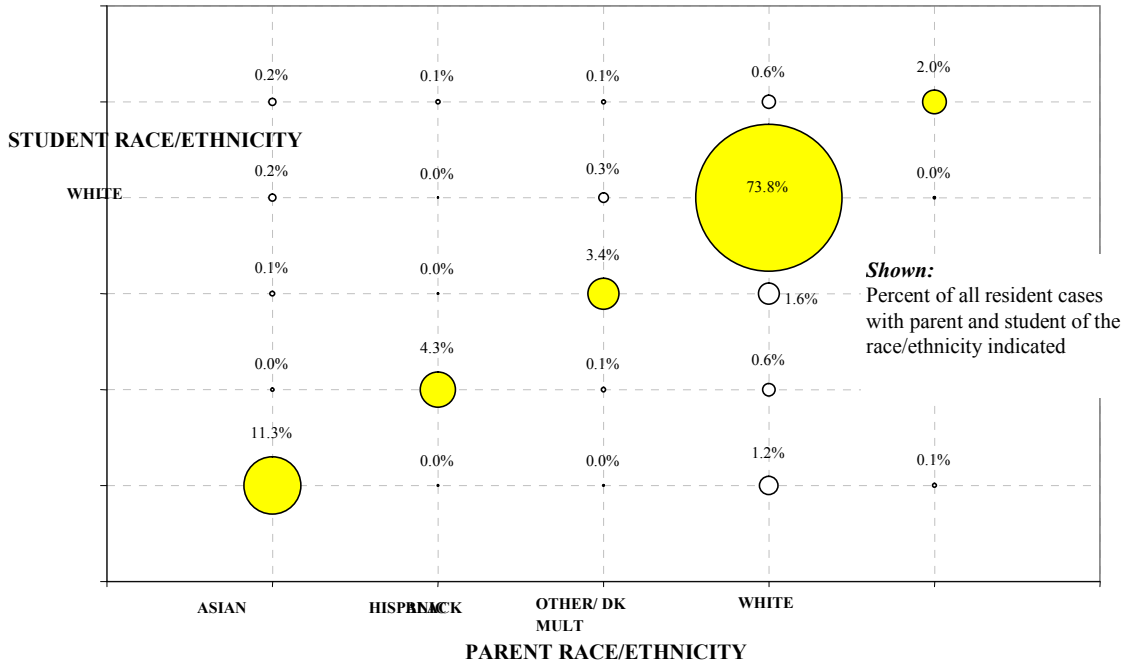
[17] No COFHE college or university has more than one percent American Indian.

[18] See National Center for Education Statistics, *The Condition of Education 2002,* Washington: U.S. Department of Education, June, 2002.

[19] One problem with the term "Asian" as a descriptor for a racial category is the confusion with the citizenship of foreign students from the continent of Asia. "Asian American" is probably a better descriptor, but the use of the word "American" with Asian and not with other categories, such as "black American" or "Hispanic American," carries different, and perhaps more offensive, connotations.

Figure 2 is a graphical presentation comparing the racial and ethnic identities of parents and children from the COFHE 2002 Parents survey. The chart depicts the (weighted) number of cases at each intersection of parent (X-axis) and student (Y-axis) race/ethnicity; the areas of the circles are proportional to the number of cases fitting the X-Y description. The vast majority of cases (95 percent) fall on the diagonal (where parent and student race/ethnicity the same) and most (74 percent) in the white-white cell. As we strive for and celebrate diversity, it is worth noting that *all* of the other combinations would comfortably fit into the white-white circle. Of special interest are the off-diagonal circles that are large relative to their corresponding diagonal ones. For example, the white parent-Hispanic student circle is almost half as large as the Hispanic-Hispanic diagonal.

FIGURE 2
Parent versus Student Race and Ethnicity



In practice, the possibility of multiple responses creates two challenges. The first concerns reporting formats: Which ethnic and racial categories should be used in tabular presentations? Which should be collapsed into subtotals or reported in a miscellaneous category? The second challenge concerns the need to devise a "bridging strategy" which connects the data collected in the multiple response format to other historical or administrative data collected in a single-response format. We address these two challenges below.

Reporting Formats. We experimented with a number of formats for reporting multiple ethnic and racial identifications. One reporting format that allows for multiple responses of race is to present the percentages of the total number of respondents who identified with each racial group regardless of whether they also identified with another group. Called "All Inclusive," the strategy includes all of the responses in a single and exhaustive display of all possible responses. Each percentage represents the portion of the population that marked each racial category. The sum of these percentages is greater than 100 percent since some people chose more than one race. Table 2 displays the results of the "All Inclusive" bridging strategy from the COFHE 2000 Alumni Survey.[20]

TABLE 2
"All Inclusive" Reporting Format

American Indian	0.8%
Asian (+ Hawaiian)	6.7%
Black	3.4%
Hispanic	2.5%
White	89.6%
Total	103.0%

A statistical strategy that sums to more than 100 percent is, however, difficult to interpret and subject to ridicule. Because graphic presentation and elementary geometry teach that the "whole must be equal to the sum of its parts," this strategy defies the intuitive understanding of "the whole" as 100 percent. 100 percent represents all of the people; 103 percent represents all of the people in all of the race categories. They are not the same due to multiple races.

When placed in the context of the politics of affirmative action in college admissions and the politics of statistical adjustments made to surveys and census counts, a strategy that totals to something greater than 100 percent is truly suspect. One can only imagine the arguments in court, a legislature, or a regulatory hearing to a distribution that totals more than 100 percent. Additionally one can only imagine comments about a presentation about racial equality where some individuals are counted more than once. While mathematically elegant, the "All Inclusive" strategy has political shortcomings.

An alternative reporting format, where the sum of the percentages of different racial groups totals 100 percent, is the "Mutually Exclusive" grouping. This representation of race data displays the percentage of individuals identifying with a single race or with any specific combination of races. There are 63 possible categories in the detailed distribution, including 6 single race groups, 15 two-race combinations, 20 three-race combinations, 15 four-race combinations, 6 five-race combination and 1 six-race combination. Table 3 displays data from the COFHE 2000 Alumni Survey.

[20] The media sometimes use a conceptually similar format with tables that include census data of race and ethnicity; the total percentage of all of the races plus Hispanics is typically greater than 100 percent.

TABLE 3
"Mutually Exclusive" Reporting Format

	Mutually Exclusive	
	Frequency	Percent
One race		
American Indian	20	0.1
Asian	1065	5.6
Black	546	2.9
Hispanic	280	1.5
White	16474	87.4
Two race		
American Indian *and* White	70	0.4
Asian (+ Hawaiian) *and* White	124	0.7
Black *and* White	42	0.2
American Indian *and* Black	13	0.1
Hispanic *and* White	149	0.8
Asian (+ Hawaiian) *and* Black	1	0.0
Hispanic *and* Black	8	0.0
American Indian *and* Asian (+ Hawaiian)	1	0.0
Hispanic *and* Asian (+ Hawaiian)	10	0.1
American Indian *and* Hispanic	3	0.0
Three race		
American Indian, Asian (+ Hawaiian) *and* White	1	0.0
American Indian, Black *and* White	18	0.1
American Indian, Hispanic *and* White	5	0.0
Asian, Black *and* White	2	0.0
Asian (+ Hawaiian), Hispanic *and* White	3	0.0
Black, Hispanic *and* White	3	0.0
Four race		
American Indian, Asian (+ Hawaiian), Black *and* White	2	0.0
American Indian, Black, Hispanic *and* White	3	0.0
Asian (+ Hawaiian), Black, Hispanic *and* White	2	0.0
Five race		
American Indian, Asian (+ Hawaiian), Black, Hispanic *and* White	7	0.0
Total	18852	100.0

The large number of multiple-race combinations is cumbersome to present and some combinations do not contain a large enough percentage of respondents to understand or differentiate the category from other groupings. Indeed the COFHE 2000 Alumni Survey had no more than 18 out of 18,852, less than one-tenth of a percent, in any category of three or more races.

An alternative to this approach is to select the combinations that are large enough to be a meaningful grouping in the data. The Aggregation Guidance by the Office of Management and Budget (OMB) present a reporting format that contains the five single race categories and the four double race combinations most frequently reported in recent studies. The format also allows for the collection of information on any multiple race combination that comprise more than one percent of the population of interest. Table 4 is an example of the reporting format.

TABLE 4
"OMB Aggregation" Reporting Format

	Frequency	Percent
American Indian	20	0.1
Asian (+ Hawaiian)	1065	5.6
Black	546	2.9
Hispanic	280	1.5
White	16474	87.4
American Indian *and* White	70	0.4
Asian *and* White	124	0.7
Black *and* White	42	0.2
American Indian *and* Black	13	0.1
Other combinations larger than 1 percent of the population	0	0.0
All other combinations of race/ethnicity	218	1.1
Total	18852	100.0

Bridging Strategies 1: Trumping Rules.[21] Statistical computations that create a single array of racial and ethnic categories and map individual categories from OMB Statistical Directive 15 to the 1997 Guidelines are known as "bridging strategies." They differ from reporting formats, which only describe an aggregation of individuals or responses for the 1997 Guidelines. Bridging strategies utilize computation rules about the meaning of the data and make assignments to categories according to a calculation rule. The greatest complication with any bridging strategy is counting multiple responses and making comparisons to a classification that had no multiple-responses.

Several different bridging strategies have been identified for creating this kind of distribution. Some involve the creation of a hierarchy among the multiple response categories, with one category dominating, or "trumping," another. The concept is not new to racial and ethnic classification. OMB Directive 15 created a hierarchy with the use of the two-question format where ethnicity typically trumps race. The nomenclature of "black, non-Hispanic" or "white, non-Hispanic" that is included in Federal government reporting implies that Hispanic trumps either white or black. However, data tables and

[21] Use of "bridge" and "trump" is less of a mixed metaphor than a bad pun.

report nomenclature never use "Asian, non-Hispanic" nor "American Indian, non-Hispanic" as names for "Asian" or "American Indian," which implies that both Asian and American Indian "trump" Hispanic.

At least four trumping rules have been proposed as plausible bridging strategies since 1997. The first (Smallest Group) assigns any person with a multiple response to the smallest group. In national data of college students and institutional data collected at COFHE schools, this decision rule would result in the following hierarchy: American Indian > Asian > Hispanic > black > white.[22] The hierarchy is sensitive to the smallest group, ensuring that those racial and ethnic classifications with the smallest numbers will be maximally counted. While mathematically elegant, the trumping rule has little resemblance to education policy.

A second trumping strategy (Largest Group) assigns responses with two or more races into the largest group. The computation using this rule is the exact opposite of the "Smallest Group" rule; larger racial categories trumping smaller ones. This decision rule reverses the hierarchy shown in the previous paragraph: white > black > Hispanic > Asian > American Indian. The logic behind it is that larger groups typically assimilate the values and culture of smaller groups rather than vice versa.

The "Largest Group" trumping strategy has the mathematical property of creating the least change to the groups that will have the largest number of cases. Because the largest groups are subject to the greatest amount of statistical analysis, "Largest Group" forces cases into the groups that are likely to be subdivided (for example, by gender) in further analysis.

Interestingly, the National Center for Education Statistics (NCES) follows a largest group trumping rule as an order for data suppression. When a small cell size precludes the release of data, which could be deduced from subtracting all remaining cells frequencies from the total, the suppression of a second cell frequency is necessary to preserve confidentiality. The decision rule for suppression of a second category that is used in the Graduation Rate Survey of IPEDS is the largest category; this rule preserves minority data.

A third alternative (Black) assigns multiple responses that include black and any other group to black, but responses with two or more racial groups other than black are assigned to the category with the fewest number of individuals. This trumping strategy recognizes practices in higher education – many of which are being challenged in the University of Michigan case before the Supreme Court. The hierarchy of racial categories reflects the emphasis on diversity among all racial and ethnic minorities in combination with the historical concern for blacks. This bridging strategy has the following hierarchy: black > American Indian > Asian > Hispanic > white.

[22] .The newly created response, "Native Hawaiian or Pacific Islander," is assigned to the old racial category of "Asian or Pacific Islander"

A fourth trumping strategy (Largest Group Other Than White) assigns multiple responses that include white with some other racial group, to the other group, but responses with two or more non-white racial groups into the group with the largest single-race count. We will discuss this bridging strategy in greater detail in the next section as it is close to the strategy that COFHE is using for reasons that will be discussed. The hierarchy is as follows: black > Hispanic > Asian > American Indian > white. The distribution of the COFHE 2000 Alumni Survey data for the four trumping strategies is displayed in Table 5.

TABLE 5
Trumping Rules for Bridging Strategies

	Smallest Group	Largest Group	Black	Largest Group Other Than White
American Indian	0.8	0.1	0.5	0.5
Asian (+ Hawaiian)	6.5	5.8	6.4	6.6
Black	3.1	3.0	3.4	3.4
Hispanic	2.4	1.5	2.3	2.3
White	87.3	89.6	87.3	87.3
Total	100.0	100.0	100.0	100.0

Bridging Strategies 2: Fractional Rules*:* Trumping strategies redistribute individuals with multiple responses to one and only one racial category. An alternative is to redistribute individuals partially according to a fraction that is derived from the data. One fractional rule (Equal Fractions) assigns all of the multiple responses from a respondent in equal fractions to each of the respondent's racial or ethnic category.[23] For example, a multiple response of white and American Indian counts as one-half in the tabulations for American Indians and one-half in the tabulations for whites. These fractions reflect how often people might identify with one group or another. Mathematically, they assign the person to groups according to the only information that the person provided.

Another fractional redistribution (Response Fractions) assigns each of the individuals with multiple responses to categories in proportion to the percentages of all responses identifying with each racial group. For any kind of statistical analysis, such as cross tabulations or regression, this is the same as weighting the multiple-races according to the percentage of responses in each racial category. Table 6 displays percentages from the COFHE 2000 Alumni Survey using fractional rules to redistribute multiple races.

[23] "Equal Fractions" treats multiple responses the same as the "All Inclusive" Reporting Format. The difference is that the former weights each response in a manner that allows the total to equal 100 percent of the respondents.

TABLE 6
Fractional Rules for Bridging Strategies

	Equal Fractions	Response Fraction
American Indian	0.4	0.1
Asian (+ Hawaiian)	6.2	5.8
Black	3.1	3.0
Hispanic	2.0	1.5
White	88.4	89.6
Total	100.0	100.0

All of the bridging strategies suggested above have inherent strengths and weaknesses. They simulate individuals' responses according to rules that the respondent either did not or would not have chosen even if given the opportunity. They make assumptions about intent where the individual was not consulted. Indeed, bridging strategies adopt procedures of data manipulation that were explicitly rejected with the creation of multiple responses.

The COFHE Bridging Strategy

The adoption of the single-question multiple-response format for a recent set of surveys of students, parents, and alumni necessitated adopting a bridging strategy for the datasets and reports produced by the Consortium. As we reviewed various alternatives, we also created a summary of the desirable characteristics of each bridging strategy:

- Succinctness. The list of summary categories should be short, which may require folding very small groups into larger ones.
- Comparability. The new summary categories should be comparable to data collected in a single-response format.
- Verisimilitude. The summary should not marginalize the "multiples" into tabular footnotes, nor overstate minority populations by assuming all "multiples" are "minorities," nor overstate the majority population by assuming that "multiples" are not minorities.

We emphasize that this process is an exercise in summarization. None of the information captured in multiple responses is lost. Users of the data can adopt alternative summary concepts or use the multiple response variables for other purposes. For example, the schema described below does not directly permit the creation of a variable for under-represented minorities (because of how American Indians are coded). As we accumulate a longer time series of data collected in this format, we will have the opportunity to revisit this issue on a regular basis.

The summary concept is based on responses to the six race/ethnicity dichotomous variables corresponding to the six racial and ethnic categories. The recoding rules are as follows:[24]

1. The classification applies only to US residents.
2. Black supercedes other identities; that is, a respondent choosing African American and *any other combination* of identities is summarized as black.
3. Hispanic supercedes identities other than black.
4. Native American supercedes identities other than black or Hispanic, except that those who identify Native American and white (only) are summarized as white. (This is discussed further below.) Because of their small number, Native Americans are often summarized in "Other".
5. Hawaiian and Pacific Islander are summarized with Asian; and Asian supercedes white.
6. Those selecting more than two categories, none of which are black or Hispanic, are summarized as "Other".
7. Those not responding to the question are summarized in "Other."

This schema contains several contestable elements and any such rubric includes choices that are somewhat arbitrary— there is no clear best way of summarizing multiple response cases. In particular, two of the "trumping" choices embedded in the summarization are worth additional comment.

First, we have chosen to allow the selection of a black identification to supercede the simultaneous choice of an Hispanic one. We are aware that others have made the opposite choice. For example, the Census two-question format implicitly allows Hispanic to trump all racial designations. Additionally, the nomenclature that is often used in government reports, i.e. "black, non-Hispanic" implies an opposite choice. Our choice reflects the historical centrality of the black-white division in the United States and current practices for minority recruitment at many colleges and universities.

Data from the 2002 COFHE Senior Survey help assess the impact of this choice. Table 7 displays the (summarized) race/ethnicity of students. About 7% of students who identified themselves as Hispanic also identified themselves as black, and vice versa. While this represents only 0.2 percent of all cases, the change does shift observed proportions. [25] If we were to use the 2000 Census instead of our own survey data, we would reach the same conclusion: only 0.2 percent of all individuals classified themselves as both black and Hispanic (only). Thus 12 percent of the population is self-categorized as "black not-Hispanic" but 12.2 percent are "Black" and 12.5 percent categorized as "Hispanic" but 12.3% as Hispanic (not black).

[24] See the SPSS code below.

[25] As with all multiples, we expect that this particular overlap will increase in the future. Where 0.2 percent of students answered yes to both Hispanic and black, only 0.04 percent of their parents did so in the 2002 COFHE Parents survey.

Table 7
Student Race Reported by Parents

	Summary	Alternative with Hispanic superceding black
Black	5.4%	5.0%
Hispanic	5.0%	5.4%

The second choice we have made is to limit American Indian identity to exclude those who also choose white. Here, the issue is the large degree of overlap. In the 2002 Parents Survey, 62 percent of parents indicating an American Indian[26] heritage also indicated that they are white. Further, there are noticeable differences between those Native Americans who select white and those who do not. For example, of those selecting American Indian *only,* 29 percent have incomes of $50,000 or less, compared to 15 percent for those who select *both* American Indian and white (white alone is 11 percent). At the other end of the income scale, about 10 percent of those identifying only as American Indian have incomes over $150,000, while 26 percent of those who also identify as white do so (the white only number is higher still, 41 percent). Bottom line: The median income of the Indian-plus-white group is 44 percent higher than the Indian-only group.

In practice, the proportion of cases who are American Indian, counted either way, is rarely is greater than one percent at any school. It is typically folded into the "other" category to ensure that the total of all racial and ethnic categories equal one hundred percent.

In sum, the COFHE bridging strategy makes assumptions about racial and ethnic categories that are consistent with higher education policy and reflect the priorities that have informed administrators for over twenty-five years. Ideally, we would present data from member institutions and show how our decision rules match with demographic data that are maintained on each campus. That work is in progress. For now, we present some comparisons for populations that are similar but not identical. For example, COFHE surveys consistently have lower than expected response rates from several minority categories. Additionally campus practices for collecting, recording, and reporting racial and ethnic data differ from one member institution to the next. Table 8 presents a distribution of several data collection efforts at COFHE.

[26] The actual term used in the questionnaires is "American Indian or Alaska Native." Although earlier COFHE surveys used the term "*Native American"* some respondents choose to (mis?)interpret this as meaning "native-born American", which of course includes the vast majority of respondents.

Table 8
Student Race: Parents Survey and Administrative Sources

	Median Institution		
	Parent Survey	2001 Fall Enrollment[1]	2001 Freshmen[2]
Asian	11.3%	12.6%	13.3%
Black	4.2%	6.0%	6.2%
Hispanic	4.5%	5.9%	6.0%
White	77.8%	67.5%	62.8%
Other/Multiple/Unknown	3.0%	6.2%	11.3%
White & Other	80.8%	73.7%	74.1%

[1]20 schools, spring 2002, Fall Enrollments (all undergraduates).
[2]21 schools, December 2001, Class entering 2001.

Note that the administrative collection systems generate more "other/multiple/unknowns" than survey research; this makes the two types of percentages slightly less comparable. Since most of the unknowns are white (even if white respondents are no more likely to avoid classification), this overstates the degree of survey bias. Adding *unknown* back to *white* is a reasonable response to this and sets a lower bound for the response bias in the survey. Even after doing this, however, the survey dataset contains 6-7 percent more white respondents than the parent survey data.

Finally, we present the SPSS code that COFHE uses for calculating race and ethnicity from a series of "yes-no" racial and ethnic variables. In the above example, American Indian has been combined with "other/multiple/unknown."

```
compute RACEC = indian + asian + black + hisp + Hawaii + white
compute RACE=6.
if (white=1) RACE=5.
if (asian=1) RACE=2.
if (hawaii=1) RACE=2.
if (indian=1) RACE=1.
if (hisp=1) RACE=4.
if (black=1) RACE=3.
if (RACEC=2 and indian=1 and white=1) RACE=5.
if (RACEC>2 and black=0 and hispanic=0) RACE=1.
if (foreign=1) RACE=7.
variable label RACE "Summary of Race and Ethnicity".
value labels RACE 1 'American Indian' 2 'Asian' 3 'Black' 4 'Hispanic' 5 'White'
6 'Other/Mult/Missing' 7 'Foreign'.
execute.
```

Where racec is a count of responses

```
COMPUTE racec = natam + asian + black + hisp + hawaii + white .
```

And foreign equals 1 if citizen=3 (citizen of another country).

If the percentage of American Indians is less than one percent, the following code will move American Indian into the "Other/Multiple/Missing"add the following line:

```
Recode RACE (1=6).
```

Discussion and Conclusion

In this paper we reviewed the "bridging strategy" that COFHE uses for recently collected data about race and ethnicity. The classification schema assumes a set of hierarchical assignments of individuals to a racial or ethic category according to principles and priorities that are common to higher education.

Nevertheless, the COFHE bridging strategy is truly work in progress. We do not have a good model for trend line comparisons because institutional data suffers from the same ambiguous classification as survey data. Furthermore, survey trend data offer little help in providing a standard for comparison because of changing identities for respondents that would normally fall within identical groups. The parent-child survey data is somewhat helpful, but parental perceptions are not always the same as childhood identities. Furthermore, adoption and inter-racial marriage reduce the utility for COFHE's existing research tools to make intergenerational comparisons. In short, the racial and ethnic classification system that is prescribed in this paper requires some external validation.

With this caveat in mind, the COFHE schema resolves several issues for researchers while also leaving some unanswered. First, the COFHE racial and ethnic classification answers questions that the National Center for Education Statistics has left unanswered. For over half a decade, the higher education community has asked the Department of Education for standards that would provide rules for handling multiple response categories that can be compared with OMB Directive 15. Researchers rely on the classification system of the Federal government to develop their own data collection, storage, and reporting systems. In the absence of this guidance, COFHE has given its member institutions a standard that will allow comparison of data from the previous era to the current research agenda.

Second, we believe that the COFHE schema, with only slight modifications, has implications for all institutions of higher education. Some schools will not want to combine "Native Hawaiian and Pacific Islander" with "Asian;" the former group is

significantly large to justify separation. Indeed, the "one-percent rule" that applies to the 1997 OMB standards as a minimum for reporting multiple-response groups seems an appropriate threshold for Native Hawaiians.

Similarly, the folding of American Indians into an "Other" classification will not be appropriate for schools where the size of this population justifies analytic attention. Additionally, the aggregation of blacks, Hispanics, and American Indians into an "under-represented minority" grouping or the aggregation of blacks, Hispanics, Asians, and American Indians into a "Minority" grouping would reduce the utility of this strategy. Of course, when one views the schema in this paper as a reporting prescription, vis-à-vis a data storage recommendation, even this decision rule results in no degradation of information.

Finally, the hierarchical decision for treating multiple responses of blacks and Hispanics does not comply with government standards. This is a problem for surveys and data collection efforts where the race and ethnicity of respondents is clearly known.

However, COFHE data often compare surveys to campus-based data collection – the precise area where NCES has not and probably will not provide guidance. While there are clear definitions of each racial and ethnic identity in IPEDS, for example, there are no clear instructions for classifying respondents that identify with more than one category. Furthermore, NCES officials have for several years felt that there is no compelling reason ever to do so.

In the absence of government standards, COFHE schools have followed the decision rule that best describes current policies and practices and helps create a rationale for future policies and practices. Black recruitment has been a priority for at least three decades for admission offices and supercedes affirmative action for other groups. Accountability suggests that the classification scheme would measure progress according to the same rules that govern policy priorities.

In conclusion, the three questions that we used to organize our thoughts about racial and ethnic classification may require other answers under different educational objectives and institutional characteristics than those found at COFHE institutions. Nevertheless, we believe they are the right questions and that our answers could be attractive to researchers that have struggled to answer them since 1997.