CollegeBoard

Identifying with More Than One Ethnic and/or Racial Group

Another Examination of the Impact on Differential Item Functioning Statistics

Anita Rawls
Xiuyuan Zhang
Amy Hendrickson

January 2016

MEASUREMENT

www.manaraa.com

College Board authors:

Anita Rawls, Associate Psychometrician

Xiuyuan Zhang, Associate Psychometrician

Amy Hendrickson, Senior Director

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools.

For further information, visit www.collegeboard.org.

Contents

Tables

Executive Summary

The classification of test-takers into ethnic and racial groups ensures individuals and groups, identified in Title VI and VII of the Civil Rights Act of 1964 and the 14th Amendment to the Constitution of the United States are protected from adverse treatment (Camilli, 2006). The United States Office of Management and Budget (OMB) suggests that respondents to ethnic and racial data collection efforts have the option to select with more than one ethnicity and/or racial group. This paper focuses on the impact on the calculations of differential item functioning (DIF) statistics when respondents can select all ethnic and racial categories that apply.

Over 5,000 participants were recruited to take a full-length SAT® or PSAT//NMSQT® pilot test. In addition to the pilot test, the test-takers also completed an 11-item survey with questions related to their past academic experiences, academic plans for the future, and background demographics. All examinees were presented with the race/ethnicity question in two ways. The first was the original question, which allowed respondents to select only one race or ethnicity. The second was the pilot question, which allowed respondents to select all ethnicities and racial groups that were applicable. Based on the responses to the two questions, the test-takers were classified into reference and focal groups in three ways. Comparisons between focal and referece groups were made based on the responses to the original question, the pilot question, and the combined responses to the original and pilot questions. The Mantel-Haenszel D-DIF and standardized P-difference index were used to evaluate differential item functioning.

Key Findings

- On average, Asian respondents had the highest scores in reading, writing, and math.

- Standardized P-difference index seems to be more sensitive than Mantel-Haenszel D-DIF statistic.

Introduction

The classification of test-takers into ethnic and racial groups ensures individuals and groups, identified in Title VI and VII of the Civil Rights Act of 1964 and the 14th Amendment to the Constitution of the United States, are protected from adverse treatment (Camilli, 2006). These groups have historically been identified as American Indian, Asian, Black, and Hispanic. The classification of individuals into protected groups is becoming increasing complex because many individuals belong to more than one group (Perez & Hirschman, 2009). Rockquemore and Brunsma (2002) suggest members of biracial groups may choose "to identify with (a) a singular identity, (b) both identities (i.e., biracial), (c) a fluctuating identity (sometimes singular and sometimes biracial), or (d) a transcendent identity (i.e., no racial label)." At any point in time, a biracial adolescent who is transitioning from childhood to adulthood may change how he or she chooses to identify. Gonzalez-Backen (2013) uses an ecological framework to explain the impact of a biracial adolescent's environment on his or her choice of identity. In particular, the author suggests that the general social climate concerning ethnicity and race, the community environment, and the child's development are direct predictors of how the child may choose to identify.

Background

As early as the 1970s, the United States Office of Management and Budget (OMB) noticed a change in how the United States population self-identifies and created statistical policy directives about how to collect ethnic and racial information (OMB 1977; Skerry, 1996). In the mid to late 1990s, the OMB commissioned several panels to research the phenomenon further and provide guidance about the best way to collect ethnic and racial data. As a result of these panels, the OMB suggested that respondents to ethnic and racial data collection efforts have the option to select more than one ethnicity and/or racial group. For the purposes of this study, ethnicity is defined as Hispanic or non-Hispanic, and race is defined as American Indian, Asian, Black, White, and two or more races. At a minimum, the selection of racial groups should include the following choices: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White (OMB, 1997; OMB, 2000). In addition to the recommendation about the ethnic and racial categories to offer respondents, the OMB also gave suggestions for analysis of such data. The OMB rules suggest that reporting of ethnic and racial data include the categories offered to respondents and a category for those who identify with "two or more races." Thus all non-Hispanic respondents who identify with more than one racial group were collapsed into one category. A caveat of the suggested analysis was that any student who identified with the Hispanic ethnicity was classified as Hispanic (whether or not he or she selected additional races).

Based on the OMB recommendations for data collection and analysis, many companies and industries that heavily rely on ethnic and racial data began to review and/or change the way in which they collect these data. The College Board noted the recommendation and began to monitor the potential changes and/or impact on their process for collection of ethnic and racial data. One such study that evaluated the impact of the change in data collection was conducted by Wendler, Feigenbaum, and Escandón (2001). This study of the change in data collection thoroughly examined how test-takers self-identify when given the opportunity to select more than one response. Based on the U.S. OMB's recommendations regarding the collection of ethnic and racial data, the Wendler et al. study piloted two new ways in which to word the ethnic and racial group identification question (See Table 1). After collecting the data for both questions, the authors explored potential shifts in categories by regions in the U.S. In addition, the study closely examined how the pilot questions would impact the differential item functioning (DIF) statistics or the ability to detect items that may be statistically biased for various groups. The researchers noted the impact to the average Mantel-Haenszel D-DIF statistics were minimal and recommended that the College Board continue to monitor the issue and to update the pilot questions to include multiple racial groups for Asians.

A similar study was conducted by Kim (2011) and examined two new pilot questions based on the findings from the previous Wendler et al. study (2001; See Table 1). The pilot questions for the Kim study compared the original single-part question to the two-part question, but it also allowed respondents to select the a new option "I do not wish to respond." Instead of taking a deep dive into understanding shifts in categories by region, the Kim study examined the standardized P-difference index. Despite the use of a different sample and a different measure of DIF statistic; the Kim study also noted that the impact on the standardized P-difference indices were minimal. The major similarities between the Wendler et al. and the Kim studies were the use of two pilot questions, the methods used to classify students, and the recommendation to continue to monitor the impact on DIF statistics.

This study represents another examination of how the change in the collection of racial/ethnic data may impact DIF statistics. The pilot question for the current study was slightly different than previous studies. The number of Hispanic categories from which the test-takers were able to identify remained the same; however, there was a change in the type of Hispanic choices. In particular, the respondents in this study were able to specifically select Cuban, Mexican, and Puerto Rican. Another minor change was to supply the test-taker with additional information to clarify the description of the Asian, Black, and White categories. A final change from previous studies was the notion that respondents were aware that the second question was for research purposes only. This study was similar to previous studies in the structure of the original question and the method for classification of the students for the DIF analysis. Given the slight variations in the wording of previous questions, this paper presents a third examination of the impact on the DIF statistics when students are allowed to identify with more than one racial or

ethnic group. This paper focuses on the impact on the calculations of DIF statistics when respondents can select all ethnic and racial categories that apply.

www.manaraa.com

Method

Assessment and Sample

Based on grade level, over 5,000 participants were recruited to take a full-length SAT or PSAT/NMSQT pilot test. The assessment in the Wendler, Feigenbaum, and Escandón (2001) and Kim (2011) studies was the SAT; this study uses data from the SAT and the PSAT/NMSQT. The PSAT/NMSQT is the Preliminary SAT/NationalMerit Scholarship Test and is typically used by students as practice for the SAT. The structure and content of the SAT and PSAT/NMSQT are similar, but the PSAT/NMSQT has fewer questions than the SAT. The exams were taken by ninth- to 12th-grade students in their local high schools and required a time commitment of at least 3 hours, Test-takers were provided with a $25 incentive for taking the test, and schools received $10 per test-taker for participation. Each exam consisted of four separately timed sections: reading, writing and language, math without a calculator and math with a calculator. For the SAT, there were 52 multiple-choice reading items, 44 multiple-choice writing items, and 58 math items. For the PSAT/NMSQT, there were 47 multiple-choice reading items, 44 multiple-choice writing items, and 48 math item (See Table 2). For all exams, the test-takers participated in up to two additional timed sections of varying subject areas (i.e., reading, writing and language, math, analysis in science or analysis in history/social studies). The data from the two additional timed sections are not evaluated in this study.

In addition to the pilot test, the test-takers also completed an 11-item survey with questions related to their past academic experiences, academic plans for the future, and background demographics. All examinees were presented with the original race/ethnicity question and the pilot race/ethnicity question. Before the second (pilot) question was the statement, "The following questions are for research purposes only. They are different from the previous 'describe yourself' question in that these questions ask you to choose as many options as you identify with. Please answer both questions about Hispanic origin and about race. For the following questions about your identity, Hispanic origins are not races."

Performance Statistics and DIF Statistics

The performance of the test-takers was examined by calculating the mean and the standard deviation of the raw criterion scores for the reference and focal groups identified below. The reference group consisted of the respondents who selected only White for the original and/or pilot question. A focal group was created for each of the following groups: American Indian, Asian, Black, Hispanic, and two or more races where appropriate. The comparisons between the reference and focal groups were calculated in three ways for this pilot study. The first set of DIF statistics classified test-takers based on their response to the original question. The second set of DIF statistics classified respondents based on how they answered the pilot question. The

third set of DIF statistics classified participants based on how they responded to the original and pilot questions. This third set is most similar to the methodology of the Wendler et. al. and Kim studies.

Focal Groups for the Original Question

In the original question, respondents were asked to select one option. The focal groups for the DIF statistics related to the original question were: American Indian, Asian, Black, and Hispanic (See Table 3). Of the groups that were included in the DIF analyses, the American Indian group had the smallest population. Participants who selected other, I do not wish to respond, more than one response, or did not respond to the question were not included in the DIF analyses.

Focal Groups for the Pilot Question

In the pilot question, participants were allowed to select one or more options. The focal groups for the DIF results related to the pilot question consisted of the following groups: American Indian only, Asian only, Black only, Hispanic, and two or more races. The "only" groups consisted of respondents who selected only one race. Respondents who selected multiple Asian racial groups (i.e., Asian and Native Hawaiian) were counted in the Asian only. Participants who selected multiple Hispanic ethnic groups (i.e., Hispanic, Cuban, Mexican, and Puerto Rican) were included in the Hispanic group. Test-takers who selected Hispanic and another race were classified as Hispanic. For example, if the respondent selected Mexican, Puerto Rican, and Black, he or she was assigned to the Hispanic group. The two or more races group combined all participants who selected more than one non-Hispanic race.

Table 4 shows the frequency of each group. The White racial group was selected most often, followed by the Hispanic ethnic group. The non-Hispanic option was selected, but non-Hispanic participants were only classified by the selected racial group. The Black and Asian racial groups were selected often. The two or more racial group was selected more than the American Indian group. There were also some test-takers who did not respond to the question.

Focal Groups for the Combined Original and Pilot Questions

When the responses from the original and pilot question were considered in combination, five focal groups were created. Respondents who selected the same non-Hispanic racial group for the original and pilot question were assigned to one of the only groups (i.e., American Indian only, Asian only, and Black only). If the participant selected one non-Hispanic category for the original question, then selected at least one distinctly different racial category for the pilot question, he or she was   assigned to the two or more racial group. Respondents who selected Hispanic on either question were assigned to the Hispanic group. Table 5 gives a pictorial description of the reference and focal groups when the responses from the original and pilot

question were combined. Tables 6 and 7 give the percentage of the sample within each category for SAT and PSAT/NMSQT, respectively.

Mantel-Haenszel D-DIF and Standardized P-difference Index

After the creation of the focal and reference groups, the DIF statistics were calculated. Based on the formulas from Dorans and Holland (1993), the Mantel-Haenszel D-DIF (MH D-DIF) statistic and the standardized P-difference index were calculated. The following formula was used for the Mantel-Haenszel D-DIF statistic

$$MH\ D - DIF = -2.35\ln[\alpha_{MH}]$$

where $\boldsymbol{\alpha_{MH}}$ is an estimate of the odds ratio. MH D-DIF values that were not statistically different from zero were classified as *A* items. Items that exceeded 1.5 in absolute value and were significantly larger than 1.0 in absolute value were classified as *C* items. The remaining values were classified as *B* items. The formula used to calculate the standardized P-difference index was:

$$STD\ P - DIF = \frac{\sum_m N_{fm} P_{fm}}{\sum_m N_{fm}} - \frac{\sum_m N_{fm} P_{rm}}{\sum_m N_{fm}}$$

where N_{fm} is the number of examinees at score point *m* in the focal group, P_{fm} is the proportion of correct responses at score point *m* in the focal group, and P_{rm} is the proportion of correct responses at score point *m* in the reference group.

Items with standardized P-difference index values between -0.05 and 0.05 were considered negligible or *low*. Items with standardized P-difference index values between -0.10 and -0.05 or between 0.05 and 0.10 were categorized as *medium*. Items with standardized P-difference index values less than or equal to -0.10 or greater than or equal to 0.10 were categorized as *high.*

It should be noted that recommended sample sizes for reference and focal groups are generally in the 100 to 300 range (Zwick, 2012). Sample sizes for several focal groups included here, notably American Indian, are smaller than recommended (See Tables 3 and 4). It is likely that DIF statistics are not calculated for groups this small in operational data. This study includes data for focal groups with at least 50 test-takers for research purposes.

Results

Performance of Test-Takers

Tables 8–11 describe the performance of SAT and PSAT/NMSQT test-takers, respectively, when classified based on their response on the original, pilot, and combined original and pilot. Tables 8 and 10 give the performance by subject area in terms of raw scores. Tables 9 and 11 describe the performance of test-takers in terms of scale scores. The section scale scores, Evidence-Based Reading and Writing (EBRW), and Math, are reported on a 200–800 scale for SAT and a 160–760 scale for PSAT/NMSQT. The Evidence-Based Reading and Writing section score is calculated using the raw scores from the reading and writing and language timed test sections. The Math score is calculated using the raw scores from the math timed test sections.

On average, Asian respondents had the highest scores in reading, writing, and math on the SAT and the PSAT/NMSQT. For the SAT, White respondents were the next highest-performing group for reading, writing, and math using the classification for the original and combined questions. However, when the respondents were categorized using the pilot question, the two or more races subgroup slightly outperformed the White respondents in writing and math. For the PSAT/NMSQT, the two or more race subgroup performed slightly lower than the Asian and White subgroups in the pilot and combined questions. When categorized using the original, pilot, and combined questions for the SAT and PSAT/NMSQT, the American Indian, Black, and Hispanic subgroups were the lowest-ranking subgroups. The Hispanic SAT and PSAT/NMSQT participants performed slightly higher than the Black subgroup in reading, writing, and math.

Classifcation of Items

Using the MH D-DIF categories (*A*, *B*, and *C*) and the standardized P-difference index categories (*low*, *medium*, and *high*), items were classified and collectively summarized in Tables 12 and 13 for SAT and PSAT/NMSQT, respectively. Regardless of which question(s) were used for classifications, most of the items were classified as having *A* or *B* MH D-DIF or *low* or *medium* standardized P-difference index. More SAT and PSAT/NMSQT items were classified as having *high* standardized P-difference indices than items classified as having *C* MH D-DIF.

When using the classifications from the original question, the American Indian subgroup had the most SAT and PSAT/NMSQT items with *high* standardized P-difference index. The Asian subgroup ranked just below the American Indian subgroup for the number of items classified with *C* MH D-DIF and with *high* standardized P-difference index for the original question. For the Hispanic subgroup classification using the original question, there was one SAT item classified as *high* standardized P-difference index and one PSAT/NMSQT item classified as having *C* MH D-DIF.

For the Black and Hispanic subgroups for the original, pilot, and combined questions, there was one SAT item classified as *high* standardized P-difference index. The Asian subgroup had the most items classified as C MH D-DIF or *high* standardized P-difference index for the pilot and combined questions. None of the statistics suggested a review of SAT items for the Hispanic and two or more races subgroups when classified using the pilot and combined questions. None of the statistics suggested a review of SAT and PSAT/NMSQT items for the two or more races subgroup when classified using the pilot and combined questions.

www.manaraa.com

Discussion

During operational testing, items classified with C MH D-DIF or with a *medium* or *high* standardized P-difference index are very carefully reviewed (Kim, 2011; Zwick, 2012). Based on the results of this study, the standardized P-difference index seems to be more sensitive than the MH D-DIF statistic. It may be beneficial to use the standardized P-difference index to recommend more items for review and to allow item writers to identify characteristics of items that produce a high standardized P-difference index. This situation of more potential false positives than false negatives seems appropriate in this context. It may be, however, that the standardized P-difference index is more sensitive to sample size. This information, although not a part of the original line of questioning for this study, was a very informative conclusion. This issue warrants further investigation.

The number of test-takers who selected American Indian only reduced to less than 50 when the test-takers were given the option to select more than one race or ethnicity. Thus, there was a gap in our ability to study the change in the American Indian subgroup. Despite this gap, it appears that the level of sensitivity for DIF detection was about the same when examinees selected only one race or ethnicity versus when they selected multiple ethnicities and/or racial subgroups. This study provides enough information for us to make a recommendation about which direction to take for analysis of DIF statistics when respondents are able to select all applicable ethnic and racial categories. However, due to study findings and limitations, caution should be taken in generalizing the results of this study.

One interesting finding was that no items for the two or more race subgroups on SAT or PSAT/NMSQT were classified with C MH D-DIF, or with a *medium* or *high* standardized P-difference index. DIF was also negligible for the Black subgroup on PSAT/NMSQT. This finding may not hold true for other samples. One limitation was that the items included in this study were previously pretested and screened for DIF, so it was challenging for us to find many items exhibiting DIF. Another limitation of our sample was that test-takers may have lacked motivation to do their best. The students received a $25 incentive, but they may not have been concerned about their level of performance. Students were aware that scores from the exam were not going to be reported. This knowledge may have also impacted how test-takers responded to the ethnic and racial questions. For example, a test-taker may have chosen two different ethnic and racial groups just because he or she was given the opportunity to respond to the question more than once. Finally, the representativeness of the sample for the target population for this exam was not evaluated in this paper; one may question the ability to achieve the same results when using a different sample of test-takers. Allowing for possible issues with generalizability, the classification of respondents must be studied with additional samples until there is some stability in the results.

The results of this current study may be used to inform future classification of focal groups for performance and DIF statistics when the way in which the College Board collects racial/ethnic information changes. To solidify classification rules for future operational examinations, the two-part question will be included in future pilot studies, along with the original single-part question, and analyses contained in this report will be repeated. The most valuable results from these future pilots will be those regarding the effect of the two-part question and of the various classification rules. Future pilot results will also inform how subgroup performance will change when the two-part question is implemented operationally.

♦ CollegeBoard

Tables

Table 1. Ethnic and Racial Questions from all Studies

Wendler, Feigenbaum, Escandón, 2001	Kim, 2011	This Study

Student Descriptive Questionnaire question:
How do you describe yourself? (Mark only one.)
American Indian, or Alaska Native
Asian, Asian American, or Pacific Islander
Black or African American
Hispanic or Latino background:
    Mexican or Mexican American
    Puerto Rican
    Other Hispanic, Latino, or Latin American
White
Other
	Original SAT question
	How do you describe yourself? (Mark only one.)
	American Indian or Alaska Native
	Asian, Asian American, or Pacific Islander
	Black or African American
	Mexican or Mexican American
	Puerto Rican
	Other Hispanic, Latino, or Latin American
	White
	Other
		Original question:
		How do you describe yourself? (Mark only one.)
		American Indian or Alaska Native
		Asian, Asian American, or Pacific Islander
		Black or African American
		Mexican or Mexican American
		Puerto Rican
		Other Hispanic, Latino, or Latin American
		White
		Other
		I do not wish to respond.

First administration question
How do you describe yourself? Mark *all* choices that apply.
American Indian or Alaska Native
Asian or Asian American
African American or Black
Hispanic or Latino background:
    Mexican or Mexican American
    Puerto Rican
    Latin American, South American, Central
    American, or other Hispanic or Latino
Native Hawaiian or other Pacific Islander
White
Other
	Version A Pilot Question
	Please check one or more of the following options that you identify with.
	American Indian or Alaska Native
	Asian or Asian American
	Native Hawaiian or Other Pacific Islander
	Black or African American
	Mexican or Mexican American
	Puerto Rican
	Other Hispanic, Latino, or Latin American
	White
	Other
	I do not wish to respond
		Pilot question:
		The following questions are for research purposes only. They are different from the previous "describe yourself" question in that these questions ask you to choose as many options as you identify with. Please answer both questions about Hispanic origin and about race. For the following questions about your identity, Hispanic origins are not races.
		
		What is your ethnicity?
		Hispanic or Latino (including Spanish origin)
		(You may mark more than one.)
		    Cuban
		    Mexican

Second Administration questions:
1. How do you describe yourself? Choose *only* one.
    Hispanic or Latino
    Not Hispanic or Latino
2. How do you describe yourself? Mark *all choices* that apply.
    American Indian or Alaska Native
    Asian or Asian American
    African American or Black
    Native Hawaiian or Other Pacific Islander
    White
    Other
	Version B Pilot Question
	1. Are you Hispanic or Latino (including Spanish other Spanish origin)?
	    Yes, Mexican or Mexican American
	    Yes, Puerto Rican
	    Yes, Other Hispanic, Latino, Latin American
	    No
	    I do not wish to respond
	2. Please check one or more of the following options that you identify with.
	    American Indian or Alaska Native
	    Asian or Asian American (including Indian Subcontinent)
	    Black or African American (including African and Afro-Caribbean)
	    Native Hawaiian, or Other Pacific Islander
	    White (including Portuguese, Brazilian, Persian, and Middle Eastern)
	    I do not wish to respond.
		Puerto Rican
		    Other Hispanic or Latino
		Not Hispanic or Latino
		
		What is your race?
		(You may mark more than one.)
		American Indian or Alaska Native
		Asian (including Indian subcontinent and Philippines origin)
		Black or African American (including African and Afro-Caribbean origin)
		Native Hawaiian or Other Pacific Islander
		White (including Middle Eastern origin)

www.manaraa.com

CollegeBoard

Table 2. Number of Items on Assessments

Subject	SAT	PSAT/NMSQT
Reading	52	47
Writing	44	44
Math	58	48
TOTAL	**154**	**139**

Table 3. Frequency of Groups Based on Original (Single-Selection) Question

Groups	SAT		PSAT/NMSQT	
	Frequency	Percentage	Frequency	Percentage
American Indian (AI)	53	0.9%	61	1.0%
Asian	501	8.6%	625	9.9%
Black	839	14.3%	880	14.0%
Mexican or Mexican American	702	12.0%	869	13.8%
Puerto Rican	50	0.9%	53	0.8%
Other Hispanic, Latino or Latin American	726	12.4%	862	13.7%
White	2,334	39.9%	2,240	35.5%
Other	162	2.8%	190	3.0%
I do not wish to respond	145	2.5%	183	2.9%
More than one response	18	0.3%	28	0.4%
Missing	323	5.5%	314	5.0%
TOTAL	**5,853**	**100.0%**	**6,305**	**100.0%**

Table 4. Frequency of Groups Based on the Pilot (Multiple-Selection) Question

Groups	SAT		PSAT/NMSQT	
	Frequency	Percentage	Frequency	Percentage
American Indian	34	0.6%	37	0.6%
Asian	463	7.9%	602	9.6%
Black or African American	777	13.3%	813	12.9%
Hispanic	328	5.6%	364	5.8%
Cuban	97	1.7%	156	2.5%
Mexican	764	13.1%	941	14.9%
Puerto Rican	67	1.1%	84	1.3%
Other Hispanic	375	6.4%	437	6.9%
Two or more Hispanic ethnicities	92	1.6%	99	1.6%
Not Hispanic	33	0.6%	45	0.7%
Native Hawaiian	4	0.1%	18	0.3%
White	2,192	37.5%	2,097	33.3%
Two or more races	229	3.9%	256	4.1%
Missing	398	6.8%	356	5.7%
TOTAL	**5,853**	**100.0%**	**6,305**	**100.0%**

Table 5. Pictorial Description of Racial/Ethnic Group Formation for Original and Pilot Questions

Pilot Question	Original question									
	Missing	American Indian	Asian	Black	Mexican or Mexican American	Puerto Rican	Other Hispanic, Latino or Latin American	White	Other	I do not wish to respond
Missing	NONE	AIO	AO	BO	HISP	HISP	HISP	WO	NONE	NONE
Hispanic	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP
Cuban	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP
Mexican	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP
Puerto Rican	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP
Other Hispanic	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP
Two or more Hispanic ethnicities	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP	HISP
Not Hispanic	NONE	AIO	AO	BO	HISP	HISP	HISP	WO	NONE	NONE
American Indian	AIO	AIO	TOMR	TOMR	HISP	HISP	HISP	TOMR	TOMR	AIO
Asian	AO	TOMR	AO	TOMR	HISP	HISP	HISP	TOMR	TOMR	AO
Black or African American	BO	TOMR	TOMR	BO	HISP	HISP	HISP	TOMR	TOMR	BO
Native Hawaiian	AO	TOMR	AO	TOMR	HISP	HISP	HISP	TOMR	TOMR	AO
White	WO	TOMR	TOMR	TOMR	HISP	HISP	HISP	WO	TOMR	WO
Two or more races	TOMR	TOMR	TOMR	TOMR	HISP	HISP	HISP	TOMR	TOMR	TOMR

AIO = American Indian Only, non-Hispanic   AO = Asian Only, non-Hispanic   BO = Black Only, non-Hispanic   HISP = Hispanic
WO = White Only, non-Hispanic   TOMR = Two or more races, non-Hispanic   NONE = Not included in DIF analysis

www.manaraa.com

CollegeBoard

Table 6. Percentage of SAT Students in each Ethnic and/or Racial Group

Pilot Question	Original question										Total
	Missing	American Indian	Asian	Black	Mexican or Mexican American	Puerto Rican	Other Hispanic, Latino or Latin American	White	Other	I do not wish to respond	
Missing	96.9%	0.0%	1.0%	1.7%	1.0%	0.0%	0.4%	1.5%	3.3%	11.0%	398
Hispanic	0.3%	5.7%	0.6%	0.6%	12.1%	0.0%	26.7%	1.1%	3.3%	4.1%	328
Cuban	0.0%	0.0%	0.2%	0.1%	0.0%	0.0%	10.2%	0.5%	2.8%	2.8%	97
Mexican	0.3%	7.5%	0.0%	0.6%	81.2%	0.0%	16.3%	1.4%	11.7%	9.0%	764
Puerto Rican	0.6%	1.9%	0.0%	0.7%	0.0%	90.0%	0.1%	0.1%	3.3%	2.8%	67
Other Hispanic	0.0%	3.8%	1.0%	1.1%	1.9%	0.0%	39.8%	1.1%	7.2%	12.4%	375
Two or more Hispanic ethnicities	0.9%	0.0%	0.0%	0.4%	3.4%	10.0%	6.2%	0.1%	2.8%	3.4%	92
Not Hispanic	0.0%	0.0%	0.8%	0.5%	0.4%	0.0%	0.0%	0.6%	2.2%	2.1%	33
American Indian	0.3%	47.2%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	1.7%	0.7%	34
Asian	0.0%	1.9%	90.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.3%	2.8%	463
Black or African American	0.3%	0.0%	0.0%	89.3%	0.0%	0.0%	0.0%	0.0%	8.3%	7.6%	777
Native Hawaiian	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	1.1%	0.0%	4
White	0.3%	1.9%	0.4%	0.1%	0.0%	0.0%	0.1%	90.4%	20.0%	28.3%	2,192
Two or more races	0.0%	30.2%	5.8%	5.0%	0.0%	0.0%	0.0%	3.0%	28.9%	13.1%	229
Total	**323**	**53**	**501**	**839**	**702**	**50**	**726**	**2,334**	**180**	**145**	**5,853**

المنارة للاستشارات

www.manaraa.com

Table 7. Percentage of PSAT/NMSQT Students in each Ethnic and/or Racial Group

Pilot Question	Original question										
	Missing	American Indian (AI)	Asian	Black	Mexican or Mexican American	Puerto Rican	Other Hispanic, Latino, or Latin American	White	Other	I do not wish to respond	Total
Missing	93.3%	1.6%	0.5%	1.7%	0.7%	0.0%	0.5%	1.0%	2.8%	2.7%	356
Hispanic	0.3%	4.9%	0.2%	0.5%	11.6%	1.9%	23.9%	1.5%	1.4%	6.0%	364
Cuban	0.0%	1.6%	0.0%	0.3%	0.0%	0.0%	13.8%	1.0%	2.8%	2.7%	156
Mexican	1.9%	3.3%	1.0%	0.5%	81.8%	0.0%	15.4%	1.6%	9.2%	13.1%	941
Puerto Rican	0.6%	1.6%	0.0%	0.6%	0.2%	84.9%	0.7%	0.3%	5.0%	3.3%	84
Other Hispanic	0.6%	0.0%	1.0%	0.9%	2.6%	0.0%	39.3%	1.5%	6.9%	5.5%	437
Two or more Hispanic ethnicities	0.3%	0.0%	0.0%	0.1%	2.6%	13.2%	6.1%	0.1%	3.2%	2.7%	99
Not Hispanic	0.0%	1.6%	0.3%	1.4%	0.1%	0.0%	0.0%	0.7%	2.8%	4.4%	45
American Indian	0.0%	49.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.5%	1.6%	37
Asian	1.0%	0.0%	89.6%	0.0%	0.0%	0.0%	0.0%	0.1%	10.1%	7.7%	602
Black or African American	1.0%	3.3%	0.3%	86.8%	0.0%	0.0%	0.0%	0.1%	9.2%	10.9%	813
Native Hawaiian	0.0%	0.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.5%	18
White	0.6%	0.0%	0.2%	0.6%	0.2%	0.0%	0.2%	89.0%	17.9%	29.0%	2,097
Two or more races	0.3%	32.8%	4.6%	6.7%	0.0%	0.0%	0.0%	3.0%	28.4%	9.8%	256
Total	**315**	**85**	**412**	**607**	**778**	**161**	**1,470**	**3746**	**247**	**142**	**6,305**

Identifying with More Than One Ethnic and/or Racial Group

www.manaraa.com

Table 8. SAT Raw Scores by Subgroup for Original, Pilot, and Combined Questions

Group	Test Section	Original			Pilot			Original and Pilot		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
American Indian	Reading	53	24.8	9.0	34	22.4	8.7	27	23.6	9.2
	Writing	53	23.6	8.0	34	20.6	8.1	27	21.8	8.2
	Math	53	24.7	10.2	34	20.8	9.1	27	21.9	9.4
Asian	Reading	501	27.9	10.3	467	28.0	10.2	465	28.0	10.2
	Writing	501	27.4	9.3	467	27.5	9.2	465	27.5	9.3
	Math	501	34.0	11.3	467	34.3	11.2	465	34.3	11.2
Black	Reading	839	20.7	8.0	777	20.5	7.9	779	20.5	7.9
	Writing	839	20.3	7.9	777	20.1	7.8	779	20.1	7.8
	Math	839	22.3	8.6	777	22.1	8.6	779	22.1	8.6
Hispanic	Reading	1,478	21.8	8.8	1,723	22.0	8.8	1,738	21.9	8.9
	Writing	1,478	20.5	8.2	1,723	20.7	8.3	1,738	20.7	8.3
	Math	1,478	23.9	8.8	1,723	24.0	8.8	1,738	23.9	8.8
Two or more races	Reading				229	26.9	9.0	302	26.0	8.8
	Writing				229	26.1	8.0	302	25.0	8.2
	Math				229	29.0	9.8	302	28.0	9.9
White	Reading	2,334	26.9	9.2	2,192	27.0	9.2	2,200	26.9	9.2
	Writing	2,334	25.8	8.4	2,192	25.9	8.4	2,200	25.9	8.4
	Math	2,334	28.6	9.6	2,192	28.8	9.6	2,200	28.8	9.5
No DIF group	Reading	648	22.2	9.2	431	20.7	9.1	342	20.5	9.0
	Writing	648	21.2	8.9	431	20.0	8.6	342	19.6	8.6
	Math	648	24.4	9.9	431	23.3	9.8	342	23.1	10.0

Table 9. SAT Scaled Scores by Subgroup for Original, Pilot, and Combined Questions

Group	Test Section	Original			Pilot			Original and Pilot		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
American Indian	EBRW	53	481	94.5	34	450	94.0	27	463	98.0
	Math	53	462	103.0	34	424	91.7	27	434	95.1
Asian	EBRW	501	523	113.2	467	524	111.9	465	525	112.9
	Math	501	555	113.8	467	558	112.8	465	558	113.3
Black	EBRW	839	438	87.6	777	436	86.2	779	435	86.4
	Math	839	440	86.3	777	438	85.8	779	438	85.6
Hispanic	EBRW	1,478	445	94.7	1,723	447	95.2	1,738	447	95.4
	Math	1,478	455	86.9	1,723	456	87.2	1,738	456	87.3
Two or more races	EBRW				229	508	95.7	302	497	95.8
	Math				229	506	97.4	302	496	98.2
White	EBRW	2,334	506	98.7	2,192	508	99.3	2,200	508	99.3
	Math	2,334	502	94.2	2,192	503	94.3	2,200	503	94.2
No DIF group	EBRW	648	452	102.7	431	436	99.8	342	433	99.8
	Math	648	460	99.0	431	449	98.7	342	448	100.8

Table 10. PSAT/NMSQT Raw Scores by Subgroup for Original, Pilot, and Combined Questions

Group	Test Section	Original			Pilot			Original and Pilot		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
American Indian	Reading	61	19.3	7.5	37	18.7	6.6	35	18.7	6.7
	Writing	61	19.9	7.9	37	17.9	7.1	35	18.4	7.0
	Math	61	15.2	5.4	37	14.5	5.9	35	14.9	5.8
Asian	Reading	625	25.3	8.1	620	25.4	8.1	598	25.4	8.1
	Writing	625	27.5	8.3	620	27.5	8.4	598	27.5	8.5
	Math	625	24.8	9.1	620	24.9	9.3	598	25.0	9.3
Black	Reading	880	17.7	6.7	813	17.5	6.6	814	17.5	6.6
	Writing	880	18.5	7.0	813	18.4	7.0	814	18.4	7.0
	Math	880	14.4	5.4	813	14.2	5.1	814	14.2	5.2
Hispanic	Reading	1,784	18.6	7.1	2,081	18.9	7.2	2,096	18.9	7.2
	Writing	1,784	19.3	7.4	2,081	19.6	7.5	2,096	19.6	7.5
	Math	1,784	15.6	5.8	2,081	15.8	6.1	2,096	15.8	6.1
Two or more races	Reading				256	22.9	8.2	358	22.4	8.3
	Writing				256	23.6	8.4	358	23.2	8.6
	Math				256	18.1	7.3	358	18.1	7.4
White	Reading	2,240	23.5	8.1	2,097	23.6	8.1	2,086	23.5	8.1
	Writing	2,240	24.8	8.2	2,097	24.9	8.2	2,086	24.9	8.2
	Math	2,240	19.2	7.4	2,097	19.3	7.4	2,086	19.3	7.4
No DIF group	Reading	715	19.5	7.8	401	17.6	6.8	318	17.6	6.9
	Writing	715	20.3	8.4	401	18.5	7.4	318	18.6	7.4
	Math	715	16.2	7.5	401	15.0	6.2	318	14.9	6.4

CollegeBoard

Table 11. PSAT/NMSQT Scaled Scores by Subgroup for Original, Pilot, and Combined Questions

Group	Test Section	Original			Pilot			Original and Pilot		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
American Indian	EBRW	61	438	89.5	37	421	80.0	35	425	78.5
	Math	61	435	69.3	37	424	76.4	35	431	75.0
Asian	EBRW	625	521	95.2	620	521	96.3	598	521	96.7
	Math	625	546	99.6	620	548	101.9	598	549	101.6
Black	EBRW	880	418	79.8	813	417	78.6	814	416	78.5
	Math	880	424	73.3	813	421	70.6	814	421	71.2
Hispanic	EBRW	1,784	429	84.1	2,081	433	86.0	2,096	432	86.0
	Math	1,784	440	76.8	2,081	442	79.7	2,096	442	79.6
Two or more races	EBRW				256	482	96.2	358	477	97.9
	Math				256	470	88.8	358	471	90.5
White	EBRW	2,240	493	94.6	2,097	494	94.9	2,086	494	94.8
	Math	2,240	485	87.7	2,097	486	87.8	2,086	485	87.5
No DIF group	EBRW	715	441	95.5	401	417	83.6	318	417	83.8
	Math	715	444	93.1	401	430	81.6	318	429	83.3

Identifying with More Than One Ethnic and/or Racial Group

www.manaraa.com

Table 12. Number of Items by DIF Category for SAT by Original, Pilot, and Combined Questions

Group		Original MH D-DIF A	B	C	Original STD P-DIF Low	Medium	High	Pilot MH D-DIF A	B	C	Pilot STD P-DIF Low	Medium	High	Original and Pilot MH D-DIF A	B	C	Original and Pilot STD P-DIF Low	Medium	High
American Indian	Reading	39	4	0	24	14	5												
	Writing	39	3	0	26	12	4												
	Math	43	3	0	26	16	4												
Asian	Reading	48	4	0	41	11	0	48	4	0	41	11	0	48	4	0	42	10	0
	Writing	34	10	0	29	15	0	32	12	0	27	16	1	33	11	0	29	14	1
	Math	45	11	2	42	11	5	47	8	3	37	16	5	46	9	3	37	17	4
Black	Reading	51	1	0	49	3	0	51	1	0	47	5	0	51	1	0	48	4	0
	Writing	42	2	0	36	8	0	42	2	0	36	8	0	42	2	0	36	8	0
	Math	54	4	0	53	4	1	54	4	0	53	4	1	54	4	0	53	4	1
Hispanic	Reading	51	1	0	43	8	1	51	1	0	45	7	0	51	1	0	41	11	0
	Writing	41	3	0	32	12	0	43	1	0	36	8	0	42	2	0	38	6	0
	Math	57	1	0	53	5	0	58	0	0	54	4	0	57	1	0	53	5	0
Two or more races	Reading							52	0	0	51	1	0	52	0	0	51	1	0
	Writing							44	0	0	42	2	0	44	0	0	43	1	0
	Math							56	2	0	52	6	0	56	2	0	53	5	0

Note: The N count for American Indian was less than 50 test-takers when categorized using the pilot question and combined original and pilot question; thus, no DIF statistics were reported.

www.manaraa.com

CollegeBoard

Table 13. Number of Items by DIF Category for PSAT/NMSQT by Original, Pilot, and Combined Questions

Group		Original						Pilot						Original and Pilot					
		MH D-DIF			STD P-DIF			MH D-DIF			STD P-DIF			MH D-DIF			STD P-DIF		
		A	B	C	Low	Medium	High	A	B	C	Low	Medium	High	A	B	C	Low	Medium	High
American Indian	Reading	44	2	0	26	18	2												
	Writing	42	2	0	29	12	3												
	Math	39	4	0	23	14	6												
Asian	Reading	45	2	0	43	4	0	45	2	0	42	5	0	45	2	0	42	5	0
	Writing	41	1	2	31	12	1	41	1	2	32	11	1	41	1	2	32	11	1
	Math	40	7	1	34	13	1	40	8	0	34	14	0	41	7	0	35	12	1
Black	Reading	47	0	0	38	9	0	47	0	0	41	6	0	47	0	0	39	8	0
	Writing	44	0	0	38	6	0	44	0	0	37	7	0	44	0	0	39	5	0
	Math	48	0	0	44	4	0	47	1	0	44	4	0	47	1	0	44	4	0
Hispanic	Reading	45	2	0	39	8	0	45	2	0	40	7	0	45	2	0	39	8	0
	Writing	42	2	1	35	9	0	43	1	0	36	8	0	42	2	0	36	8	0
	Math	48	0	0	47	1	0	48	0	0	47	1	0	48	0	0	46	2	0
Two or more races	Reading							46	1	0	41	6	0	46	1	0	42	5	0
	Writing							43	1	0	43	1	0	44	0	0	40	4	0
	Math							48	0	0	46	2	0	48	0	0	46	2	0

Note: The N count for American Indian was less than 50 test-takers when categorized using the pilot question and combined original and pilot question; thus no DIF statistics were reported.

الــمـنـارة للاستشارات

www.manaraa.com

Bibliography/References

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational Measurment Fourth Edition* (p. 221–256). Phoenix, AZ: American Council on Education/Onyx Press Series on Higher Education.

Dorans, N.J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland and H. Wainer (Eds.) *Differential Item functioning* (p. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Gonzales-Backen, M. A. (2013). An application of ecological theory to ethnic identity formation among biethnic adolescents. *Family Relations, 62(1),* 92–108.

Kim, Y. R. (2011). *Differential item functioning analysis for multi-race/ethnicity subgroups.* Unpublished manuscript. New York, NY: College Board.

Perez, A. D., & Hirschman, C.. (2009). The changing racial and ethnic composition of the U.S. population: Emerging American Identities. *Population and Development Review, 35(1),* 1– 51.

Rockquemore, K. A., & Brunsma, D. L. (2002). Socially embedded Identities: Theories, typologies and processes of racial identity among black/white biracials. *The Sociology Quarterly, 43(3),* 335-36.

Skerry, P. (1996). *Many American dilemmas: The statistical politics of counting by race and ethnicity.* Washington, DC: The Brookings Institute. Available from: http://www.brookings.edu/research/articles/1996/06/summer-immigration-skerry

United States Office of Management and Budget (1977). *Race and ethnic standards for Federal statistics and administrative reporting* (Statistical Policy Directive No. 15). Washington, DC: U.S. Office of Management and Budget. Available from: http://wonder.cdc.gov/wonder/help/populations/bridged-race/Directive15.html

United States Office of Management and Budget (1997). *Revisions to the standards for the classification of Federal data on race and ethnicity.* (Federal Register Notice, October 30, 1997).
Washington, DC: U.S. Office of Management and Budget. Available from: http://www.whitehouse.gov/omb/fedreg_1997standards/

United States 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). Washington, DC: U.S. Office of Management and Budget.

Wendler, C., Feigenbaum, M., & Escandón, M. (2001). *Defining group membership: The impact of multiple versus single ethnic/racial categories on testing practices.* (College Board Research Note No. 12). New York, NY: The College Board.

Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement (ETS Research Report No. RR-12-08). Princeton, NJ: ETS.

00078_014

www.manaraa.com