Two Steps Forward, One Step Back:

Race/Ethnicity and Student Achievement in Education Policy Research

Bruce D. Baker*

Christine Keller-Wolff

Lisa Wolf-Wendel

Department of Teaching and Leadership

202 Bailey Hall

The University of Kansas

Lawrence, KS 66045

(785) 864-9844

bdbaker@ukans.edu

*Corresponding Author

6-May-00

Biographical Statements

Bruce D. Baker is an assistant professor in the Department of Teaching and Leadership at the University of Kansas, Lawrence. Dr. Baker specializes in Education Policy, Economics and quantitative methods. His previous work in Educational Policy includes an article on the use of vouchers to equalize opportunities for low-income gifted children (with Craig E. Richards in July, 1998) and a methodological critique of the Politics of the Production Function (January/March, 1999).

Christine Keller-Wolff is a doctoral candidate in the Higher Education Program in the Department of Teaching and Leadership at the University of Kansas, Lawrence. She works as a research analyst in the Office of Institutional Research and Planning at the University of Kansas.

Lisa Wolf-Wendel is an associate professor in the Department of Teaching and Leadership at the University of Kansas, Lawrence. Dr. Wolf-Wendel specializes in equity issues affecting Higher Education, with particular emphasis on
special focus colleges.

Abstract

The goal of this study is to bring the discussion of ethnic heterogeneity and the racial/ethnic classification of students for research purposes into the education policy arena. We focus in particular on the relationship between race and ethnicity and academic achievement. We demonstrate the heterogeneity of academic performance in reading and math between subgroups of Hispanic and Asian/Pacific Island students using the National Educational Longitudinal Study of 1988 (NELS '88). In the case of both the Hispanic and Asian/Pacific Island aggregate groups we find substantial, though not always statistically significant, academic performance differences among ethnic subgroups, with a range of math performance among Hispanic subgroups of 10.7 points (mean score = 34.4) between Cuban and Puerto Rican students and a range of math performance among Asian/Pacific Island students of 15.3 points (mean score = 41.0) between West Asian and Pacific Island students.

Key Words: Race, Ethnicity, Academic Achievement

Historical Perspective

The relationship between race/ethnicity and academic achievement has been studied extensively for years (Gibson & Ogbu, 1991; Neisser, 1986; Sleeter & Grant, 1993; Weinberg, 1977). Over the past four decades, there has been a gradual evolution toward refining research definitions of racial/ethnic groups, and more specifically, toward disaggregating ethnic subgroups to the extent possible. This evolution, however, has not always progressed consistently toward greater disaggregation. In the 1960's and 1970's, most researchers used the term "minority students" to describe students who were not White. Because the majority of these "minority students" were Black students, researchers typically classified students according to a Black/White dichotomy (Sellars & Weis, 1997). In the early 1980's, education researchers began to identify Hispanic students as a separate group (Marin & Marin, 1991). Students of other national origins, including Chinese, Filipino or Native Hawaiian students were typically combined with White students (Dawson, 1987) or excluded from analyses altogether. American Indian and Alaskan Native students were largely absent from the education literature before 1990, except as discussed in a limited number of federal or state-sponsored reports (Reyhner, 1989).

In the last two decades, increased sensitivity of researchers to racial and ethnic diversity in public schools has led to frequent use of five standard groups: White, Black, Asian/Pacific Island, Hispanic and American Indian/Alaskan Native. Most recently, researchers interested in ethnic diversity have begun to explore differences within these groups, focusing on specific ethnic subgroups, including Puerto Rican students as members of the larger Hispanic group (Donato & Wojtkiewicz, 1996), and Vietnamese students in the Asian/Pacific Island context (Wilson, 1994). These studies extensively discuss heterogeneity among Hispanic students and Asian/Pacific Island students, raising concerns regarding potentially misleading stereotypes attached to aggregate groups, including the perception of Asian/Pacific Island students as a "model minority" and Hispanic students as "low achievers" (Lai, 1990; Nieto, 1991; Sellars & Weis, 1997; Suzuki, 1995; Vigil, 1997). While examples of disaggregation of Black and American Indian/Alaskan Native students in education research are less common, others have noted that they are no less important (Gibson & Ogbu, 1991; Nieto, 1991; Sellars & Weis, 1997).

Race and Ethnicity in Education Policy Research

With a few notable exceptions (Klein et al., 1997) recent education policy research rarely addresses racial and ethnic issues as a central question, but rarely discounts them altogether either. Most recent policy research in education ascribes to the five group classification system (Asian, White, Black, Hispanic, Native American), taking one step back along the disaggregation continuum (Klein et. al, 1997; Raudenbush, Fotiu, & Cheong, 1998; Gamoran et al., 1997). Still others in policy research choose to take two steps back, continuing to apply dichotomous specifications (Taylor, 1997; Wang, 1998), often de-emphasizing the aggregations by tucking them away as "well accepted" control variables to be discussed as necessary among policy implications.

Interestingly, despite de-emphasis of race in many studies, race-related conclusions and policy implications often surface in discussions of findings. A less than random sampling of recent (1997 to present) issues of a popular education policy journal yielded the following race-related findings:

- Using data from the National Educational Longitudinal Study of 1988 (NELS '88), researchers found that Hispanic students do less well than White students in spatial-mechanical reasoning, comparable to White students in quantitative science, and less well than White students in basic knowledge reasoning on the multiple choice components of National Assessment of Educational Progress (NAEP) math and science tests (Hamilton, 1998).
- Using NAEP data, Asian students were found to be more likely than White students, and Hispanic students less likely than White students to attend a school where algebra is offered and where math teachers majored in math (Raudenbush, Fotiu & Cheong, 1998).
- Using the High School and Beyond (HSB) dataset, researchers concluded that Black men experience 13% lower earnings than White men, and Black women have 22% lower earnings than White women. There were no significant differences between Latinos and Whites for either gender (Miller, 1998)

Not coincidentally, each of these studies uses popular national datasets developed by the National Center for Education Statistics (NCES). These datasets contain varied degrees of specificity regarding race/ethnicity, with NELS '88 disaggregating four subgroups of Hispanic students and ten of Asian/Pacific Island students, but most other datasets containing only the standard five group classification. Thus, one might argue that in many cases researchers are constrained by the available data. We find it particularly intriguing, however, that some researchers choose not to take advantage of disaggregated classifications where available, especially given the presumption of within group heterogeneity discussed in research outside of education policy. One rationale for aggregating, even with large datasets like NELS '88, is the issue of subgroup sample size to be addressed in more detail with respect to our own analysis. Trading likelihood of statistical significance for the meaningfulness of the measure, however, is certainly questionable. For example, what meaningful inferences can be drawn from Hamilton's finding that the Hispanic aggregate group of NELS '88 students does not perform well on spatial-mechanical reasoning activities where the aggregate group consists of distinctly different subgroups of Cuban, Mexican, Puerto Rican and Other Hispanic students?

Purpose of the Study

The goal of this study is to bring the discussion of ethnic heterogeneity and the racial/ethnic classification of students for research purposes into the education policy arena. While we expect students to be heterogeneous in a variety of ways, we focus in particular on their academic achievement, as academic achievement differences are of critical interest in education policy research and have not been comprehensively documented. We demonstrate the heterogeneity of academic performance in reading and math between subgroups of Hispanic and Asian/Pacific Island students using NELS '88. We choose NELS '88 for is prevalence in policy research in recent years.

Our analyses are designed to test the extent to which commonly applied ethnic/racial classifications (White, Black, Asian/Pacific islander, Hispanic and Native American) appropriately represent their constituent subgroups with respect to academic achievement and the extent to which disaggregating adds explanatory value to statistical models of student achievement. Hispanic subgroups investigated include Mexican[1], Cuban, Puerto Rican and other Hispanic, while Asian subgroups include Chinese, Filipino, Japanese, Korean, Southeast Asian, Pacific islander, South Asian, West Asian, Middle Eastern and Other Asian. Data limitations inhibit our ability to explore heterogeneity among White, Black and Native American students.

While the primary emphasis of this article is the relationship between aggregate and disaggregate racial and ethnic classifications and math and reading performance measures, we must also consider major confounding factors. Socioeconomic status (SES) differences between both aggregate and disaggregate racial and ethnic groups and the relationship between these SES differences and student performance are well known (Coleman, 1966). Similarly, language proficiency has been shown to display a strong relationship to student performance and is likely related to aggregate and disaggregate racial and ethnic classifications.[2] While there are many other likely covariates with racial classification and student performance, we limit our analyses to these two additional factors, SES and language proficiency, so as to focus more clearly on differences between aggregated and disaggregated analyses.

Methods

Data

This study uses data from the public use version of the NELS '88, a nationally representative, multipurpose study of the educational status of approximately 25,000 students collected by the National Center for Education Statistics (NCES). A sample of 14,596 8th grade students was used in this study. Exclusions were made for missing or incomplete data on variables of interest. As previously noted, NELS ’88 is among the few large data sets that contain detailed subgroup classifications for both Hispanic and Asian/Pacific Island students. And, despite relatively small subgroup sample sizes, NELS ’88 provides the best available opportunity to explore differences among disaggregated groups.

In order to insure adequate representation, NELS ’88 involved a complex stratified sampling procedure. In particular, NELS '88 oversampled private schools and schools with high enrollments of Asian/Pacific Island students and Hispanic students. As a result, student level weights provided within the dataset must be applied to all analyses to correct for generalizability against the population of U.S. schools and students. In addition, the stratified sampling method employed by NCES results in nesting of students into groups, or schools, such that statistical measures including standard errors tend to be more variable than if the sampling were simply random. Consideration is given to NELS ’88 design effects when discussion statistical significance of findings.[3]

There are a variety of problematic issues regarding our research with NELS '88. We argue, however, that ignoring the uniqueness of disaggregate ethnic groups, or simply accepting the aggregate as meaningful either as a variable of interest or a control to be tucked into the background is at least equally problematic. One issue regarding the analyses herein, and perhaps a reason why NELS '88 has apparently not previously been used to demonstrate our main point is that the 14,596 students being investigated are unevenly distributed by racial and ethnic characteristics into nested groups in approximately 1000 schools. Thus, it becomes difficult to discriminate between student level effects on student performance as related to the variables of interest and peer group effects on student performance. One common resolution to this problem is to group (school) mean-center each analysis.[4] This approach, however, is problematic when the variable(s) of interest, in this case RACE or racial subgroup, is not randomly distributed across organizations.[5] For example, mean centering of a predominantly Black, lower SES organization would create inflated estimates of performance due to depression of the organization level mean by exogenous but correlated factors (RACE and in turn, SES).

An additional problem with the data being investigated is the relatively small sample sizes of the various racial subgroups within organizations, which excludes the possibility of developing meaningful, reliable separate within school models. Thus, for simplicity, but recognizing the shortcomings of the method, all regressions reported herein are simple single level, OLS (Ordinary Least Squares) regressions.[6]

Comparison of Mean SES and Mean Performance

Preliminary analyses involved determining means and standard errors for the socio-economic status composite variable and for 8th grade math and reading IRT (item response theory) estimated number correct for students by aggregate racial classification and disaggregate subgroup. While student level weights (BYQWT) were used to determine means for purposes of generalizability, reported standard errors for the means are understated due to the complex sampling design. The purpose of this analysis is merely to provide a cursory descriptive overview of the SES and performance characteristics of the aggregate and disaggregate groups.

Are there significant differences between subgroups?

The first set of regression analyses were designed to compare the academic achievement differences among the five standard aggregate groups in math and reading. In each case, regression models were constructed where the math or reading achievement measure was the dependent variable, race or subgroup classifications were the independent variables of interest, and controls were included for socioeconomic status and language proficiency.[7] Subsequently, comparable analyses were performed to assess the academic achievement differences between the disaggregated groups of Hispanic and Asian/Pacific Island students.[8]

What does disaggregation contribute to modeling achievement?

The final analysis involves additively replacing aggregate Hispanic then Asian/Pacific Island classifications with disaggregated classifications in separate analyses, using the full sample (*n* = 14,596). The objective of this analysis is to determine whether including disaggregated subgroups in a larger model of student achievement adds to the explanatory power of the model in addition to revealing performance differences among subgroups. We argue, however, that even if disaggregation does contribute significantly to improved model fit, the fact that disaggregation reveals differences between subgroups remains pertinent. We begin this set of analyses at the same point as our previous set of analyses, with a regression of the aggregate groups on the student achievement measures. Subsequently, we “unfold” so-to-speak the aggregate groups into the four Hispanic subgroups, followed by the ten Asian/Pacific Island subgroups.[9]

Results

Comparison of Mean SES and Mean Performance

Table 1 displays the means and standard errors of the socioeconomic status composite and math and reading scores for the aggregate race groups. There are no surprises in the analysis of the aggregate groups, with White and Asian students having generally higher socioeconomic status composite values and higher math and reading performance scores.

Insert Table 1

Table 2 displays the mean SES composite values and math and reading scores for Asian/Pacific Island subgroups. This cursory analysis suggests that some subgroups may stand out as different both with respect to their SES and academic performance. For example, Chinese students appear to have relatively low average SES with relatively high average math performance while Korean students have both high SES and high math performance. Presuming low SES to deflate performance we might expect the math performance of Chinese students to be even higher when adjusted for SES differences and that of Korean students to be lower, hence the need to control for SES in subsequent analyses. Pacific Island students appear to have both low SES and low performance relative to other subgroups.

Insert Table 2

Table 3 provides means and standard errors for SES and performance variables for the Hispanic subgroups. Among the Hispanic subgroups, Cuban students appear to stand out in terms of their relatively high SES and performance compared to other Hispanic subgroups.

Are there significant differences between subgroups?

Table 4 displays the results of the first set of regression models with math scores as the dependent variable, and Table 5 displays the comparable analysis of Reading performance. In the first models, where aggregate race classifications are used, there are again no surprises. SES effects are strong and positive, Asian students outperform White students in math, but not in reading, and all other groups (Black, Hispanic and Native American) perform poorly in both math and reading achievement relative to White students.

As noted previously, statistical significance is overstated in OLS analysis of NELS'88 data due to the assumption of random sampling. Design effects reported in the NELS'88 User's Manual may be used to correct, or upwardly adjust standard errors, and are generally considered conservative for adjusting standard errors of regression coefficients. Standard error adjustments are made by multiplying the simple random standard error times the Mean Root Design Effect. [10] Mean Root Design Effects reported for student performance scores in reading and math are 2.284 and 2.379 respectively.[11] In the case of the aggregate groups, performance differences generally remain statistically significant, even at the $p<.01$ level.

Insert Table 4

Insert Table 5

Looking at the within group analysis of the Hispanic subgroups, we find that Cuban students significantly outperform their Mexican counterparts in math, while Puerto Rican students perform slightly less well than their Mexican counterparts in math, though neither is significant when adjusted for design effects. No significant differences appear among Hispanic subgroups for reading performance. Among the Asian/Pacific Island subgroups, Pacific Island students perform particularly poor relative to Chinese students in Math, even when adjusted for design effects. Pacific Island students perform slightly better, but still significantly lower than Chinese students in reading. Filipino students perform less well than their Chinese peers in math, but outperform their Chinese peers in reading.

An interesting feature of this analysis is that language proficiency plays a significant role in differentiating performance among Asian subgroups while SES differences do not. Note that despite the substantially reduced sample size, the within group analysis of the Asian/Pacific Island category yields a much higher r-squared than either the aggregate model or the within group analysis of the Hispanic category. Altogether the within group analyses suggest that there are some subgroups that may not be appropriately represented by their common aggregate classification.

What does disaggregation contribute to modeling achievement?

Tables 6 and 7 display the results of the regression analyses where Hispanic and Asian/Pacific Island subgroups were unfolded within the context of the larger models of student math (Table 6) and reading (Table 7) performance. The tables begin by displaying estimates for the racial aggregate groups only, subsequently adding the SES and LEP controls, then substituting the Hispanic subgroups for the Hispanic aggregate classification and finally substituting the Asian/Pacific Island subgroups for their aggregate classification. As with previous analyses, the coefficients on the aggregate classifications reveal no surprises.

Given the larger sample size relative to the within group analyses, and the change in baseline comparison group to White students for all comparisons, several additional coefficients appear significant. Only a few coefficients, however, remain significant when adjusted for design effects. The relative positions of the subgroups remain consistent with the previous analysis. Unfolding the Hispanic aggregate group into its subgroups seems to add no explanatory power to either the math or reading performance model. Consistent with the within group results, we find that Mexican[12], Puerto Rican and Other Hispanic students perform significantly less well than their White peers in math and reading, though less so in reading. Cuban students perform comparably to their White peers.

Unfolding the Asian/Pacific Island aggregate group seems to provide a slight advantage to the explanatory power of the model of math achievement, as expressed in terms of adjusted r-squared. Beyond the slight improvement in model fit we find that Chinese students outperform their White peers by over 6 (15.5%) points on average in mathematics, while Pacific Island students perform more than 6 points below their White peers. The Asian/Pacific Island aggregate coefficient suggests that this group of students, on average, outperforms their White peers in mathematics by 2.35 points. Only Korean students possess a coefficient, though non-significant, that is comparable to this value, with math performance differences for other subgroups as high as 8.45 (21.9%) points (West Asian) above White means and as low as 6.83 (Pacific Island) below White means, calling into question the meaningfulness of the aggregate group coefficient. All significant coefficients for Asian subgroups, however, are negated when conservatively adjusted for design effects.

Conclusions and Implications

The analyses herein, though not yielding overwhelming statistical significance, do raise questions about the meaningfulness of aggregate racial classifications of students of diverse ethnic and cultural origins. For example, it is difficult to lend credence to a coefficient that suggests that Hispanic students perform, on average, 3.85 points below their White peers in math, when more refined analysis suggests that Cuban students perform 3.80 points above their White peers, and Mexican and Puerto Rican students perform 3.57 and 6.88 points below their White peers in math, respectively. Similarly it is difficult to forward the perception of the aggregate Asian/Pacific Island student group as a "model minority" on the basis that they outperform their White peers in math by an average of 2.35 points, when the point differential between disaggregated groups is as high as 15.28 points (Pacific Island students 6.83 points below White students and West Asian students 8.45 points above White students).

While greater disaggregation can be a useful step in developing more meaningful classifications of students, our analyses also indicate the importance of including related variables such as English language proficiency and socioeconomic status in explaining academic achievement. Statistically, when these variables are used in addition to the disaggregated race/ethnicity variables, the ability to predict student academic achievement improves. Conceptually, including such variables allows us to develop a more complete picture of the student, beyond skin color or point of geographic origin of their ancestors. But, even including these variables, over 90 percent of the variance in academic achievement remains unexplained, suggesting that our limited models of student background fall well short of comprehensively explaining academic achievement. Generational status and the differences between voluntary and involuntary migration are among the additional delineations that have been discussed as useful descriptors of student background in recent literature (Fuller, 1994; Kim, 1997; Marin, 1991; Gibson & Ogbu, 1991; Trueba, 1992).

In policy literature we often speak of the subtle distinction between statistical significance and policy significance. Such discussions are usually confined to finding reasons to discuss coefficients that do not quite meet the stringent requirements of statistical significance ($p < .05$), but appear important to policy. The analyses herein take this discussion in a slightly different, but related direction, in that we can typically meet the statistical significance requirement by aggregating students to achieve sufficient sample sizes. In fact, many of the aggregations that occur in policy data appear to be done for just this purpose. For example, why would a researcher using NELS '88 choose to derive a control variable of "percent minority peers," where the minority classification consists of the sum of Black, Hispanic and Native American students(Taylor, 1997, p. 92)? Would Asian students not be equally qualified for this category, based purely on a probabilistic definition? Perhaps including a group perceived to outperform their peers in the minority group would put the expected negative, statistically significant coefficient at risk. If this is the underlying rationale, why not exclude Cuban students and replace them with Pacific Island students to solidify the statistical significance? While tucked deep in the background of this study, the implication of this variable, as validated by its significant coefficient, is that students who attend school with Black, Hispanic and Native American students are likely to do less well academically than those who do not. The policy significance of this variable, however, is highly questionable and the social implications somewhat disturbing.[13]

Where race and ethnicity are brought nearer the surface in the policy literature even greater caution is warranted. For example, in a study of standardized tests, Wang (1998) investigated whether or not the opportunity to study the topics represented in tests influenced students' test scores. Wang found that the aggregate group of Black and Hispanic students (combined as a single category) received far lower levels of exposure, coverage and quality of instruction compared to their White and Asian peers (combined as a single category). Wang's findings leave the reader to infer that policymakers should take appropriate actions to insure that the aggregate group of Black and Hispanic students receive more comprehensive coverage of material covered on standardized tests, or that standardized tests must be more specifically tailored to these aggregate groups. Similarly, Raudenbush, Fotiu and Cheong (1998) paint a picture of vast disparity of mathematics opportunities for Black, Hispanic and Native American students relative to White and Asian students, implying that policymakers and school personnel should take action to design racially targeted policies to remedy these disparities. Our own analyses, however, would suggest that these targets are less well defined than some may presume.

Our intent is not to suggest that taking one or more steps forward along the disaggregation continuum is a panacea for understanding the relationship between student background and academic performance. It may be only slightly more rational to design an academic intervention policy for Mexican students as a specific subgroup than to design an intervention policy for the aggregate of Hispanic students. However, the more crude our specifications are, the less efficient and effective our policy efforts will be. For example, designing intervention policies, such as curriculum or assessment modification that focus on Hispanic students as an aggregate group because of their apparent disadvantage, but not on Asian students because of their apparent advantage will be inefficient because many Hispanic students are unlikely to need the assistance offered by such a policy and ineffective by disadvantaging those Asian students who do not "live up to" their aggregate group statistical profile.

We conclude with a plea to those involved in the design and development of the increasingly prevalent national datasets. These datasets, including NELS ’88, TIMSS (Third International Math and Science Study) and the most recent *Early Childhood Longitudinal Study, Kindergarten Class of 1998 – 99* are fast becoming the primary source of reliable and comprehensive data for education policy researchers across the country, and throughout the world. While international analyses raise entirely different sets of issues, datasets focused on our own culturally diverse education system in the United States should provide researchers the opportunity to paint as rich a portrait of student background as is technically feasible. NELS ’88 in particular, has helped us to take two steps forward. In the future, we hope not to be asked to take one step back.

References

Carey, N. & Farris, E. (1996). Racial and Ethnic Classifications Used by Public School. Statistical Analysis Report. (NCES-96-092) National Center for Education Statistics: Washington, DC.

Coleman, J. et al. (1966). Equality of Educational Opportunity. U.S. Department of Health Education and Welfare: Washington, DC.

Cooper, B.S., Cilo, M.R., Baker, B.D. (2000) Bridging the Gap between School and College: An Analysis of K-16 Education in New York City. Paper Presented at The Annual Meeting of the American Educational Research Association. New Orleans, LA.

Dawson, M. (1987). Minority student performance: Is the Montessori magnet school effective? ERIC document 309881.

Donato, K.M. & Wojtkiewicz, R.A. (1996). The educational achievement of U.S. Puerto Ricans. New England Journal of Public Policy, 11 (2), 99-111.

Fuller, E.J. (1994). Trust as the basis for urban school reform and as an explanation of the variability in involuntary minority academic achievement. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Gamoran, A., Porter, A.C., Smithson, J., White, P. (1997) Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth. Educational Evaluation and Policy Analysis. 19 (4) 325-338.

Gibson, M.A. & Ogbu, J.U.(1991). Minority Status and Schooling: A Comparative Study of Immigrant and Involuntary Minorities. New York: Garland Publishing.

Hamilton, L.S. (1998) Gender Differences on High School Science Achievement Tests: Do Format and Content Matter? *Educational Evaluation and Policy Analysis* 20 (3) 179-195.

Kim, H. (1997) Diversity Among Asian American High School Students. Educational Testing Services, Policy Information Center. Princeton, NJ.

Klein, S., et al. (1997). Gender and racial/ethnic differences on performance assessments in science. Educational Evaluation and Policy Analysis, 19(2), 83-98.

Lai, M.K. et al. (1990). Reading and written expression performance of ten Asian/Pacific Islander ethnic groups on the eighth grade California Assessment Program. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Marin, G. & Marin, B.V. (1991). Research with Hispanic Populations. Applied Social Science Research Method Series, Volume 23. Sage Publications: Newbury Park, CA.

Miller, S.R. (1998) Shortcut: High School Grades as a Signal of Human Capital. Educational Evaluation and Policy Analysis 20 (4) 299-311.

Neisser, U. (1986). The School Achievement of Minority Children: New Perspectives. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Nieto, S. (1991). Affirming Diversity: The Sociopolitical Context of Multicultural Education. Longman: New York.

Peng, S.S., Wright, D., Hill, S.T. (1995). Understanding Racial-Ethnic Differences in Secondary School Science and Mathematics Achievement. (NCES 95-710). US Department of Education: Washington, DC.

Raudenbush, S.W., Fotiu, R.P., & Cheong, Y.F. (1998) Inequality of Access to Educational Resources: A National Report Card for Eighth-Grade Math. Educational Evaluation and Policy Analysis 20 (4) 253-267.

Reyhner, J. (1989). Changes in American Indian Education: A Historical Retrospective for Educators in the United States. The Office of Education Research and Improvement: Washington, DC.

Sellars, M. & Weis, L. (Eds.) (1997). Beyond Black and White: New Faces and Voices in US Schools. Albany: State University of New York.

Sleeter, C. & Grant, C. (1993). Making Choices for Multicultural Education: Five Approaches to Race, Class, and Gender, 2nd Edition. Glencoe/MacMillan McGraw Hill.

Suzuki, B.H. (1995). Education and the socialization of Asian Americans: A revisionist analysis of the “model minority” thesis. In The Asian American Educational Experience: A Source Book for Teachers and Students. D.T. Nakanishi & T.Y. Nishida (Eds.).Routledge: New York, 113-132.

Taylor, C. (1997) Does Money Matter? An Empirical Study Introducing Resource Costs and Student Needs to Education Production Function Analysis. In William Fowler (Ed.) pp. 75 – 97. Developments in School Finance: 1997. Washington, DC: National Center for Education Statistics.

Trueba, H.T. (1992). Many groups, one people: The meaning and significance of multicultural education in modern America. Bilingual Research Journal, 16 (3-4), 91-116.

United States Department of Education. Third International Math and Science Study *(TIMSS).* Washington, DC: National Center for Education Statistics.

United States Department of Education. National Educational Longitudinal Study of 1988 (NELS '88): Third Follow-Up. Washington, DC: National Center for Education Statistics.

Vigil, J.D. (1997). Personas Mexicanas: Chicano High Schoolers in a Changing Los Angeles. Case Studies in Cultural Anthropology. Hartcourt Brace College Publishers: Fort Worth, TX.

Wang, Jia (1998) Opportunity to Learn: The Impacts and Policy Implications. Educational Evaluation and Policy Analysis 20 (3) 137-156.

Weinburg, M. (1977). Minority students: A research appraisal. National Institute of Education, Washington DC, 396, ERIC document 137483.

Wilson, C. (1994). Promoting an image: A Vietnamese success strategy through history. Opinion paper, ERIC document 392862.

Table 1.
Descriptive Statistics by Aggregate Racial Classification

	Socio-Economic Status			8th Grade Reading Performance		8th Grade Math Performance	
Aggregate Group	N	Mean	SE	Mean	SE	Mean	SE
Asian/Pacific Islander	979	0.12	0.09	29.8	0.52	41.0	0.55
Hispanic	2169	-0.50	0.09	27.5	0.43	34.4	0.43
Black	1645	-0.46	0.06	25.7	0.47	31.5	0.46
White	9273	0.01	0.01	29.8	0.15	38.6	0.16
Native American	530	-0.31	0.03	24.5	0.50	31.2	0.51

Table 2.
Descriptive Statistics by Asian, Pacific Island Disaggregate Group

Asian/Pacific Island Subgroup	Socio-Economic Status			8^{th} Grade Reading Performance		8^{th} Grade Math Performance	
	N	Mean	SE	Mean	SE	Mean	SE
Chinese	181	-0.01	0.07	29.0	0.94	44.9	1.05
Filipino	196	0.24	0.05	31.9	1.40	39.8	1.38
Japanese	58	0.41	0.07	31.6	1.91	44.2	2.07
Korean	107	1.11	0.82	31.7	1.07	45.9	1.30
Southeast Asian	177	-0.47	0.06	28.4	1.22	40.5	1.26
Pacific Islander	65	-0.21	0.09	23.9	2.24	31.6	2.10
South Asian	65	0.52	0.10	31.3	1.05	45.9	1.77
West Asian	23	-0.20	0.17	34.7	4.39	46.9	4.18
Middle Eastern	34	0.47	0.14	26.0	1.58	37.9	2.21
Other Asian	57	0.04	0.10	30.9	2.62	39.1	2.61

Table 3.
Descriptive Statistics by Hispanic Disaggregate Group

Hispanic Subgroup		Socio-Economic Status		8th Grade Reading Performance		8th Grade Math Performance	
	N	Mean	SE	Mean	SE	Mean	SE
Mexican	1571	-0.56	0.12	27.8	0.52	34.5	0.52
Cuban	35	0.24	0.13	33.4	4.05	42.6	3.82
Puerto Rican	148	-0.54	0.05	26.8	1.48	31.2	1.37
Other Hispanic	387	-0.32	0.04	27.2	0.89	34.8	0.95

Table 4
Math Performance by Aggregate and Disaggregate Groups

	All by Race (n = 14,596)			Hispanic by Subgroup (n = 2,169)			Asian by Subgroup (n = 979)		
	Estimate		SE	Estimate		SE	Estimate		SE
White	-	-	-						
Black	-6.78	***	0.43						
Native American	-7.10	***	0.69						
Hispanic	-3.85	***	0.43						
Mexican				-		-			
Cuban				8.31	**	3.82			
Puerto Rican				-3.12	*	1.87			
Other				0.56		1.13			
Asian/Pacific Isl.	2.35	***	0.70						
Chinese							-	-	-
Filipino							-3.36	*	1.73
Japanese							0.80		2.58
Korean							2.26		2.25
Southeast Asian							-2.16		1.92
Pacific Islander							-10.96	***	2.22
South Asian							2.72		2.43
West Asian							3.83		3.07
Middle Eastern							-5.64	*	2.96
Other Asian							-4.04	*	2.37
SES	0.79	***	0.07	0.07		0.10	0.28		0.21
LEP	-0.07		0.16	1.16	*	0.68	-3.38	***	0.93
	R^2= .037 Adj. R^2= .037			R^2= .005 Adj. R^2= .003			R^2= .065 Adj. R^2= .055		

*p<.10, **p<.05, ***p<.01

Table 5
Reading Performance by Aggregate and Disaggregate Groups

	All by Race (n = 14,596)			Hispanic by Subgroup (n = 2,169)		Asian by Subgroup (n = 979)		
	Estimate		SE	Estimate	SE	Estimate		SE
White	-	-	-					
Black	-3.86	***	0.41					
Native American	-5.13	***	0.67					
Hispanic	-2.01	***	0.41					
Mexican				-	-			
Cuban				5.93	3.82			
Puerto Rican				-0.78	1.88			
Other				-0.36	1.13			
Asian/Pacific Isl.	-0.10		0.68					
Chinese						-	-	-
Filipino						3.27	**	1.66
Japanese						2.72		2.48
Korean						2.70		2.16
Southeast Asian						0.12		1.84
Pacific Islander						-4.21	**	2.13
South Asian						2.66		2.34
West Asian						6.10	**	2.95
Middle Eastern						-2.87		2.84
Other Asian						2.25		2.28
SES	0.52	***	0.07	0.06	0.10	0.22		0.20
LEP	-0.07		0.15	1.03	0.68	-2.70	***	0.89
	R^2= .015 Adj. R^2= .014			R^2= .003 Adj. R^2= .000		R^2= .036 Adj. R^2= .025		

*p<.10, **p<.05, ***p<.01

Table 6.
Math Performance of Aggregate and Disaggregate Groups (as additive blocks)

	Race Only			Race, SES & LEP			Hispanic Disaggregate			API Disaggregate		
	Estimate		SE	Estimate		SE	Estimate		SE	Estimate		SE
White	-		-	-		-	-		-	-		-
Black	-7.16	***	0.43	-6.78	***	0.43	-6.73	***	0.43	-6.77	***	0.43
Native American	-7.36	***	0.69	-7.10	***	0.69	-7.06	***	0.69	-7.10	***	0.69
Hispanic	-4.25	***	0.43	-3.85	***	0.43				-3.84	***	0.43
Mexican							-3.57	***	0.49			
Cuban							3.80		3.46			
Puerto Rican							-6.88	***	1.66			
Other							-3.46	***	0.94			
Asian/Pacific Isl.	2.44	***	0.71	2.35	***	0.70	2.39	***	0.70			
Chinese										6.28	***	1.79
Filipino										1.04		1.45
Japanese										5.29	*	2.82
Korean										6.39	***	2.31
Southeast Asian										2.27		1.79
Pacific Islander										-6.83	***	2.30
South Asian										6.94	***	2.62
West Asian										8.45	**	3.53
Middle Eastern										-1.11		3.73
Other Asian										0.44		2.53
SES				0.79	***	0.07	0.79	***	0.07	0.79	***	0.07
LEP				-0.07		0.16	-0.07		0.16	-0.07		0.15
	R^2= .030			R^2= .037			R^2= .037			R^2= .040		
	Adj. R^2= .029			Adj. R^2= .037			Adj. R^2= .037			Adj. R^2= .039		

*p<.10, **p<.05, ***p<.01

Table 7.
Reading Performance of Aggregate and Disaggregate Groups (as additive blocks)

	Race Only			Race, SES & LEP			Hispanic Disaggregate			API Disaggregate		
	Estimate		SE	Estimate		SE	Estimate		SE	Estimate		SE
White	-		-	-		-	-		-	-		-
Black	-4.11	***	0.42	-3.86	***	0.41	-3.83	***	0.42	-3.86	***	0.42
Native American	-5.30	***	0.67	-5.13	***	0.67	-5.10	***	0.67	-5.12	***	0.67
Hispanic	-2.28	***	0.41	-2.01	***	0.41				-2.01	***	0.41
Mexican							-1.73	***	0.47			
Cuban							3.52		3.36			
Puerto Rican							-2.68	*	1.61			
Other							-2.42	***	0.91			
Asian/Pacific Isl.	-0.04		0.69	-0.10		0.68	-0.07		0.68			
Chinese										-0.79		1.74
Filipino										1.99		1.41
Japanese										1.53		2.75
Korean										1.31		2.25
Southeast Asian										-1.21		1.75
Pacific Islander										-5.75	**	2.23
South Asian										1.27		2.54
West Asian										4.96		3.43
Middle Eastern										-4.02		3.28
Other Asian										1.02		2.46
SES				0.52	***	0.07	0.52	***	0.07	0.51	***	0.07
LEP				-0.07		0.15	-0.07		0.15	-0.70		0.15
	R^2= .011			R^2= .015			R^2= .014			R^2= .015		
	Adj. R^2= .011			Adj. R^2= .014			Adj. R^2= .014			Adj. R^2= .014		

*p<.10, **p<.05, ***p<.01

APPENDIX A: VARIABLES AND SAMPLE DESCRIPTION

Sample:
N = 14,596

8th Grade: BYSES

BYSES was constructed using the following parent questionnaire data: father's education level, mother's education level, father's occupation, mother's occupation, and family income (data coming from BYP30, BYP31, BYP34B, BYP37B, and BYP80).

Binary indicator for Limited English Proficiency (for 8th grade): BYLEP

Race Indicator: BYS31A: (Re-coded Binary - White = 0, Black, API, Hispanic, American Indian/Alaskan Native)

Asian Subgroup Indicator: BYS31B (Re-coded Binary - Chinese, Filipino, Japanese, Korean, Southeast Asian, Pacific Island, South Asian, West Asian, Middle Eastern, Other)

Hispanic Subgroup Indicator: BYS31C (Re-coded Binary - Cuban, Mexican, Puerto Rican, Other)

8th Grade Achievement Test Results (IRT scores) BY2XMIRR (Math), BY2XRIRR (Reading)

For information on the standardized outcome measures consult Appendix H of the Second Follow-Up: Student Component Data File User's Manual.

Endnotes

[1] Throughout this study we refer to disaggregated subgroups of students according to the descriptors included in NELS '88: Mexican, Cuban, Puerto Rican, Chinese, Filipino, Korean etc. However, we note that it may be more appropriate to refer to the students as Cuban American, Filipino American and so on, as all students in our sample resided and attended school in United States in 1988.

[2] See, for example, Cooper, Cilo and Baker (2000)

[3] For a concise discussion of this feature of the data-set, see *Responses to Frequently Asked Questions Regarding NELS:88,* http://nces.ed.gov/surveys/nelsque.html. for a more detailed discussion, see Data File User's Manual, BY-F1 Student Component, Section 3.6.1

[4] The group mean-centered OLS regression analysis takes the form: $Y_{ij} - Y_{\cdot j} = \beta_w(X_{ij} - X_{\cdot j}) + r_{ij}$, where Y_{ij} is the outcome measure Y, perhaps math achievement, for student *i* in the j^{th} school, $Y_{\cdot j}$ is the mean performance for students in school *j*, X_{ij} is the independent measure of interest, perhaps SES, for student *i* in the j^{th} school, β_w is the coefficient of the pooled-within-organization relationship between X and Y and r_{ij} is the residual variance. Adding binary variables for racial group membership to the right hand side of the equation creates comparisons of achievement, scaled against within class means, by race controlled for SES, scaled against within class means. Bryk and Raudenbush (1992, p. 117-123) provide a detailed discussion of methods for estimating person-level effects (p. 117) and disentangling person level and compositional effects (p. 121).

[5] A GLM (General Linear Model) procedure indicated an F-statistic of 6.09 (R-squared = 0.42) using Race as a predictor of School Assignment (School ID). This finding is consistent with our basic understanding of de facto ethnic and SES clustering within and between school attendance boundaries.

[6] While these caveats may be statistically troublesome to some, consider the possible shortcomings of a hierarchical approach, using aggregate classifications applied to a similar sample. For one, with 15,000 students across 1000 schools, we would, on average have within school samples of n = 15 students. With approximately 7% of our students as Asian/Pacific Islanders, we may or may not, have one such student selected as representative of the Asian/Pacific Islander population for a given school. That student may be Korean or Chinese or perhaps even Pacific Islander. Given expected differences between these groups, this selection will likely affect the within school position of the aggregate group representative. While this aspect of discrimination is statistically desirable, it is

confounded by the fact that this student's score in turn affects his/her own position by significantly affecting the within school mean of the group of only 15 students.

[7] The first model can be expressed as:

$$ACH = \beta_0 + \beta_1 SES + \beta_2 LEP + \beta_3 BLACK + \beta_4 NATIVE + \beta_5 HISPANIC + \beta_6 API + \varepsilon \qquad (1)$$

Where ACH is achievement in either math or reading, SES is the socioeconomic status composite, LEP is the limited English proficiency indicator (1 = yes, 0 = no), ε is the error term, and racial groups are binary classifications, with White students as the baseline comparison group. API refers to Asian/Pacific Islander students. For the aggregate analysis, n = 14,596.

[8] For example, the equation for testing differences among Hispanic subgroups may be specified:

$$ACH = \beta_0 + \beta_1 SES + \beta_2 LEP + \beta_3 CUBAN + \beta_4 PR + \beta_5 OTHER + \varepsilon \qquad (2)$$

Where n = 2,169 and Mexican students serve as the baseline for comparison (PR = Puerto Rican). The equation for assessing differences among Asian/Pacific Island subgroups takes the same form, with Chinese student serving as the baseline for comparison and n = 979.

[9] The equation including disaggregation of the Hispanic subgroups is specified:

$$ACH = \beta_0 + \beta_1 SES + \beta_2 LEP + \beta_3 BLACK + \beta_4 NATIVE + \beta_5 MEXICAN + \beta_6 CUBAN + \beta_7 PR + \beta_8 OTHER + \beta_9 API + \varepsilon \qquad (3)$$

In the above equation, n = 14,596 and White students' performance serves as the baseline. We then similarly unfolded the Asian/Pacific Island group into its subgroups while reverting to the aggregate Hispanic classification in order to individually test the usefulness of disaggregating the Asian/Pacific Island group versus disaggregating the Hispanic group.

[10] See BY-F1: Student Component Data File User's Manual, p. 57.

[11] Note, however, that these are actually the design effects for the first follow-up and not the base year IRT math and science scores, as these are the only ones reported in the NELS '88 manual. The first follow-up (10^{th} grade) sample involved even more complex stratified sampling with intentional oversampling than the base year, to achieve a freshening of the sample. As a result, these design effects are highly conservative.

[12] Significant even when conservatively adjusted for design effects.

[13] We would like to note that we could have chosen any one of a number of studies to make this example, but we were most familiar with this study in particular. No offense is intended to Dr. Taylor, whose study presents major methodological advancements in the production function literature.