DOCUMENT RESUME

ED 390 927 TM 024 375

AUTHOR Lane, Suzanne; And Others
TITLE An Examination of the Performance Gains of Culturally and Linguistically Diverse Students on a Mathematics Performance Assessment within the QUASAR Project.
SPONS AGENCY Ford Foundation, New York, N.Y.
PUB DATE Mar 95
CONTRACT 890-0572
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Achievement Tests; Bilingual Education; Black Students; Cultural Differences; Economically Disadvantaged; Educational Change; *Elementary School Students; English; Ethnic Groups; Hispanic Americans; Intermediate Grades; Junior High Schools; *Junior High School Students; Low Income Groups; *Mathematics; Mathematics Achievement; Middle Schools; Performance; Poverty; Spanish; White Students
IDENTIFIERS Language Minorities; *Linguistic Pluralism; Middle School Students; *QUASAR Project (Mathematics Education); Reform Efforts

ABSTRACT

The performance of students from different racial or ethnic subgroups and of students receiving bilingual (Spanish and English) or monolingual (English only) instruction in mathematics was studied using students from schools in the QUASAR (Qualitative Understanding Amplifying Student Achievement and Reasoning) project, a mathematics education reform project supporting innovative instruction for middle school students in economically disadvantaged communities. Patterns of ethnic distribution vary across QUASAR sites, and linguistic diversity was found at many QUASAR sites. Data were from administration of the QUASAR Cognitive Assessment Instrument (QCAI) to approximately 1,000 students in grades 6, 7, and 8 at 2 QUASAR project schools, one of which provided bilingual classes for its predominantly Latino population. Results indicate that the instructional programs of both schools provided similar educational opportunities in mathematics, with parallel gains for African American and Caucasian students. Results also indicate that high quality mathematics instruction can be made available to students in bilingual classes. Evidence also supports the validity of the QCAI. (Contains 5 figures and 40 references.) (SLD)

An Examination of the Performance Gains of Culturally and Linguistically Diverse Students on a Mathematics Performance Assessment Within the QUASAR Project

Suzanne Lane, Edward A. Silver, and Ning Wang

University of Pittsburgh

Running Head: ETHNIC AND LINGUISTIC SUBGROUP PERFORMANCE GAINS

Preparation of this paper was supported by a grant from the Ford Foundation (grant number 890-0572) for the QUASAR project. Any opinions expressed herein are those of the authors and do not necessarily represent the views of the Ford Foundation. This paper was presented in the symposium, "Examining Equity Issues Embedded Within Mathematics Instructional Reform: Findings From the QUASAR Project" at the annual meeting of the American Educational Research Association , San Francisco, CA, April 1995.

An Examination of the Performance Gains of Culturally and Linguistically Diverse Students on a Mathematics Performance Assessment Within the QUASAR Project

Currently, considerable effort is being devoted to the reform of precollege education in many academic subject areas. In the area of mathematics, for example, reports from the National Academy of Sciences (National Research Council, 1989) and the National Council of Teachers of Mathematics (1989, 1991), have captured the attention of many educational practitioners and policy makers. The reports specify for mathematics education a set of goals and principles, usually referred to as standards, and they provide descriptions of desired mathematical proficiency, with respect to reasoning, problem solving, communication, and conceptual understanding. Moreover, the reports also indicate the expectation that mathematical proficiency should and can be attained by all students (Silver, 1994).

Despite the current optimism that all students can learn mathematics, it is the case that a pattern of achievement differences in mathematics and many other areas has been found for students in racial or ethnic minority subgroups (O'Connor, 1989). In his review of findings related to mathematics achievement by racial and ethnic subgroups on the National Assessment of Educational Progress (NAEP), Secada (1992) concluded:

> The general picture of racial and/or ethnic disparities in mathematics achievement that comes from the NAEP data is that Whites perform much better in mathematics than do Hispanics who, in turn, achieve slightly better than do African Americans. (p. 628)

Although the size of the performance gap between racial or ethnic subgroups narrowed during the 1970s and 1980s (Jones, 1984), gains appear to have been due primarily to increased performance on the parts of the assessment that measure basic knowledge and skills; far less change has been detected for tasks assessing the more complex forms of knowledge and proficiency that are emphasized in the current reform.

Moreover, a substantial gap still remains in the average mathematics performance on NAEP by members of racial or ethnic subgroups (Mullis, Dossey, Cambell, Gentile, O'Sullivan, & Latham, 1994). And Secada (1992) also notes several studies that have found a fairly direct relationship between language proficiency and performance on mathematics achievement tests. For example, De Avila (1988) found significant correlations between English language proficiency and CTBS mathematics achievement test scores for fourth, fifth, and sixth grade students.

In addition to findings that indicate a mathematics achievement gap between racial or ethnic subgroups and between groups differing in language proficiency, other findings suggest that the gap often increases as students progress through school. As Secada (1992) notes in his summary of research in this area, ".. achievement disparities, which are great to begin with, increase over time as students grow older" (p. 628). Support for this claim of a widening gap over time and grade level is supported not only by the NAEP data at grades 4, 8 and 12, but also by some other studies using cross-sectional analyses. For example, Gross (1988) reported that the achievement gap between White or Asian students and non-White or non-Asian students attending elementary schools in a large suburban district were small in grade 1 but progressively larger for each succeeding grade level.

In general, the findings related to achievement gaps in mathematics and other academic subjects have been obtained through the use of tests utilizing multiple-choice item formats. Some proponents of educational reform have argued that the use of alternate forms of assessment, usually referred to as "authentic" or performance assessments, could yield different patterns of results. This argument is often tied to a notion of assessment-driven instruction. According to this view, the use of performance assessments aimed at more complex types of knowledge and proficiency is likely to lead teachers to alter their instruction, so as to promote good performance by their students (e.g., Resnick & Resnick, 1992). Increased instructional attention to higher-level cognitive goals would then equalize opportunity to perform well, thereby leading to a narrowing of any existing achievement gap. Another reason for optimism that the use of performance assessments will lead to narrowing of the achievement gap is that these tasks could be more accessible to diverse populations of students. In contrast to decontextualized, multiple-choice test items, performance assessment tasks can allow for diverse approaches and solutions, thereby "tapping a wide range of talents, a variety of life experiences, and multiple ways of knowing" (Darling-Hammond, 1995, p. 99).[1]

Despite the optimism of advocates of performance assessments that differences in performance on these tasks among ethnic, racial, and linguistic subgroups would be narrower than those observed on multiple-choice tests, early evidence suggests that the performance differences are about the same regardless of item type (Baker, O'Neil, & Linn, 1991; Dunbar, 1987; Dunbar, Koretz, & Hoover, 1991; Feinberg, 1990; Linn, Baker, & Dunbar, 1991). For example, Linn et al. (1991) indicated that score differences for African-American and Caucasian students on written essays on the NAEP were about the same size as those found on multiple-choice reading tests. Of course, these results are not surprising, since the existence of performance differences is considered by many to be largely a consequence of unequal access to quality curriculum and instruction (Barr & Dreeben, 1983; College Entrance Examination Board, 1985; Darling-Hammond & Snyder, 1991; Oakes, 1990). According to this view, it cannot be expected that the form of assessment will have a major impact on the quality of performance unless the quality of instruction improves for minority students, since the persistent pattern of performance

---

[1] It is worth noting that not everyone agrees with these arguments. For example, several commentators have pointed to weaknesses in the logical necessity and/or practical feasibility of the argument for assessment-driven instructional reform (e.g., Silver, 1992; Winfield, 1995). Also, because evaluation of performance assessments typically involves subjective judgment, several commentators have pointed to their susceptibility to scoring bias that could disadvantage students from diverse backgrounds (e.g., O'Connor, 1989; Winfield, 1995). Moreover, it can be argued that performance assessments, which tend to involve more use of language than do multiple-choice test items are likely to be especially problematic for students whose native language is not the one used in the assessment. For example, in their review of research on mathematics learning by bilingual learners, Wong Fillmore and Valadez (1986) discuss the special challenges that tasks given in verbal formats present to non-native speakers of English when they are learning mathematics in English.

differences discussed above is seen as related to limitations of conventional instruction and disparities in access to high-quality instructional opportunities (Neill, 1995).

Given that alternate forms of assessment can be used not only as indicators of proficiency but also as indicators of instructional opportunity, especially with respect to high-level cognitive goals, they have been linked closely to current efforts to reform mathematics instruction (Silver, 1992). If students are provided with new types of mathematics instruction designed to promote high-level goals, such as reasoning and problem solving, then it should be possible to detect the impact of this instruction on students through the use of mathematics performance assessment. It should then also be possible to examine the extent to which racial and/or ethnic subgroups or subgroups reveiving instruction in different languages receive similar opportunities to develop proficiency with more complex forms of knowing and doing mathematics.

In this paper we present the results of an analysis of the performance of students from different racial or ethnic subgroups and of students receiving bilingual (Spanish and English) or monolingual (English only) instruction in mathematics. The schools attended by these students are participating in the QUASAR project, which is a mathematics education reform project that supports the design and implementation of innovative mathematics instructional programs for middle school students in economically disadvantaged communities (Silver, 1993). QUASAR schools are located in urban school districts, and they serve a culturally and linguistically diverse set of students. Aggregated across all QUASAR schools, about half the students are African-American, about one-third are Latino, and about one-eighth are Caucasian. The patterns of ethnic distribution of the student population vary across sites: two schools serve predominantly African-American students, two serve primarily Latino students, and the other two schools have student populations that are internally somewhat more ethnically diverse. Linguistic diversity is also found in many QUASAR schools. In fact, most schools serve large subgroups of students for whom English is not the primary language spoken at home; in two schools, the majority of students live in homes where English is not the primary language of discourse.

In QUASAR, the mathematics teachers and school administrators at each school work with each other and with "resource partners" -- who are typically mathematics educators from a local university -- to enhance the school's mathematics instructional program with an emphasis on mathematical understanding thinking, reasoning, and problem solving. Each site team operates independently in designing and implementing its own plan for curriculum, staff development, and other features of its instructional program. Thus, there is diversity across the schools with respect to the particular curriculum used and the forms of instructional support provided for students. Amidst this diversity, however, there is a common commitment to providing instruction that encourages development of students' conceptual understanding of mathematical ideas and their capacity for mathematical thinking, reasoning and problem solving.

There is evidence that this commitment is actually being carried out in the instruction provided in QUASAR classrooms. In particular, classroom observation data collected systematically over the course of the project suggests that QUASAR instruction promotes thinking, reasoning, problem solving, and communication to a much greater extent than is found in conventional mathematics classrooms. In an analysis of a representative sample of nearly 150 instructional tasks used in project classrooms over three years, Stein, Grover, and Henningsen (in press) found that about three-fourths of the instructional episodes involved mathematical tasks intended to provoke students to engage in conceptual understanding, reasoning or problem solving. These tasks encouraged students to use rather sophisticated mathematical thinking and reasoning -- either connecting procedures to underlying concepts and meaning or tackling complex mathematical problems in novel ways. Only about 20% of the tasks were set up to involve computation or memorization of information without some overt connection to developing understanding. These proportions stand in stark contrast to Stodolsky's

(1988) findings in which 97% of the mathematics classes she observed dealt with low-level cognitive objectives. Not onlyis there evidence of new forms of instruc on in QUASAR classrooms, but there is also evidence that students are deriving benefits in the intended direction.

The extent to which the instruction in QUASAR classrooms has beneficial effects on students has been examined by measuring changes in students' mathematical performance over time. Since high-level cognitive outcomes are intended to be a special focus of instruction at QUASAR sites, the primary source of evidence concerning the extent to which students' mathematical thinking and reasoning is affected by instruction in QUASAR classrooms is the QUASAR Cognitive Assessment Instrument (QCAI), which was developed specifically to assess students' mathematical understanding, problem solving, reasoning, and communication (Lane 1993: Silver & Lane, 1993). An analysis of QCAI results obtained from first three project years at the four original QUASAR schools found clear evidence that students developed an increased capacity for mathematical reasoning, problem solving and communication during that time period (Lane & Silver, 1994). Evidence of changes in students' mathematical understanding, thinking and reasoning over time came from an aggregation of holistic judgments of student performance on a QCAI tasks administered in all three project years and at all three grade levels. In particular, the number of students providing responses judged to be at the two highest score levels more than doubled (from 18% to 40%) between Fall 1990 and Spring 1993. Further evidence was obtained from a detailed examination of changes in students' mathematical knowledge, in their use of appropriate strategies, and in the quality of their mathematical justifications on a selected set of QCAI tasks which have been publicly released. This analysis provides clear evidence that students in project schools are developing deeper mathematical understandings and an increased capacity for complex mathematical thinking and reasoning over time.

Since the overall findings of previous analyses suggest that the mathematics instruction in classrooms in QUASAR schools embodies many of the features advocated in the current reform, and since there is also evidence that students are gaining in mathematical proficiency on the kinds of tasks valued in the current reform, it is reasonable to ask wheher all students appear to be benefiting from the instruction. In fact, since the project is attempting to provide quality mathematics instruction to **all** students in middle schools that serve students living in poverty, regardless of whether they are African American, Caucasian, or Latino, it is important to examine the patterns of proficiency and progress among ethnic and linguistic subgroups of students. Therefore, the purpose of this study is to examine the patterns of mathematics performance and changes in performance over time for various ethnic and linguistic subgroups within selected QUASAR Schools.

METHOD

Sample

Given the diversity in curricula and other aspects of instructional emphasis across project schools, and given variance in the ethnic and linguistic composition of student populations across sites, the examination of proficiency and progress for students in various ethnic and linguistic subgroups is complicated. Since two schools serve predominantly Latino students and two other schools serve predominantly African American students, differences detected in the progress made by students in ethnic and linguistic subgroups would be confounded with instructional program differences among the sites, if students from all six project schools were inclded in the analysis. A more appropriate approach is to examine changes in student proficiency for differing ethnic and linguistic subgroups within a school. If instruction is provided in an equitable manner at the school, all students should be expected to have similar opportunities to learn. If performance differences are found for ethnic and linguistic subgroups within a school, these differences will need to be carefully interpreted to determine whether they suggest differential opportunity to learn or differential capacity to benefit from mathematics instruction oriented toward high-level cognitive goals.

The data for this study were from the administration of the QCAI to 6th, 7th, and 8th grade students attending two of the QUASAR school sites in the Fall and Spring of the 1990-1991, 1991-1992, 1992-1993, and 1993-1994 school years. These two school sites were chosen because these schools had achieved the greatest progress on the QCAI during the time period considered in this analysis, and because there were at least two different cultural and/or linguistic subgroups within each school. The student population at School A consisted of approximately 135 students at each of the three grade levels, with approximately three times as many Caucasian as African American students. At each grade level at School B there were approximately 200 students, of whom about 60% were African American students, about 20% were Latino, and about 20% were Caucasian. At School B, most of the Latino students were enrolled in bilingual classes. Thus, School A afforded an opportunity to examine patterns in performance changes over time for Caucasian and African American students, and School B provided a setting in which to examin changes in performance for students in English-speaking classes and students in bilingual classes, as well as for the examination of performance gains for Caucasian and African American students. The sample consisted of approximately equal numbers of male and female students. The precise number of students in each subgroup at each of the schools at each point in time is reported in the Results section.

Assessment Instrument

At the time the project was initiated in 1989, there were no existing assessment instruments for middle school mathematics that were aligned with key features of the reform-oriented conceptualization of mathematical proficiency (e.g., problem solving, reasoning, communicating) and that had sufficient reliability and validity evidence to support their use. Therefore, the project developed and validated its own assessment instrument, which was called the QUASAR Cognitive Assessment Instrument (QCAI)[2].

The QCAI is designed to measure student outcomes and growth in mathematics, and to help evaluate attainment of the goals of the mathematical instructional programs (Lane, 1993). The QCAI consists of a set of open-ended tasks that assess students' mathematical understanding, problem solving, reasoning, and communication. Throughout the development process, steps were taken to ensure that the QCAI assesses students' knowledge of a broad range of mathematical content, understanding of mathematical concepts and their interrelationships, and capacity to use high-level thinking and reasoning processes to solve complex mathematical tasks (NCTM, 1989). Figure 1 provides examples of QCAI tasks[3].

Insert Figure 1 about here

There are two versions of the QCAI: one appropriate for the 6th and 7th grade levels and another appropriate for the 8th grade level. Each version of the QCAI consists of 36 open-ended tasks, which are distributed into four forms, each containing nine tasks (Lane, Stone, Ankenmann, & Liu, 1994). The 8th grade version of the QCAI consists of approximately half the tasks that are in the 6th/7th grade version, and the remaining tasks are unique to the 8th grade version. This allows for longitudinal analyses across 6th, 7th, and 8th grade students. Although the forms in each of the two versions are not

[2] A number of papers have provided evidence for the validity of the QCAI in terms of content quality and cogntive complexity (Lane, 1993; Lane, Parke, & Moskal, 1992; Magone, Cai, Silver, & Wang, 1994; Magone, Wang, Cai, & Lane, 1994), generalizability of the QCAI (Lane, Stone, Ankenmann, & Liu, 1994; Lane, Liu, Ankenmann, & Stone, in press), and gender-related differential item functioning (Wang & Lane, 1994).

[3] The QCAI is secure. The decision to keep the QCAI secure was based in part on the belief that evidence obtained from the assessment regarding student performance and program accountability would be more credible if the tasks were kept secure and also in part on the impractical and technical demands of developing a large number of tasks each year for assessing change in student performance. The items appearing in Figure 1 are tasks that appeared on the QCAI during the period 1990-1993 but that are now released and longer part of the current versions of the QCAI.

considered to be parallel, the tasks were distributed systematically across the forms to help ensure that the forms were as similar a possible with regard to content, processes, modes of representation (text, picture, graph, tables), context, and difficulty.

Because two of the project sites serve predominantly Latino students, a Spanish bilingual version of the QCAI was developed. The Spanish bilingual version of the QCAI presents the English and Spanish version of the task on adjacent pages so that students have the option to read the task in Spanish and/or English and to respond in either language.[4]

Administration of the QCAI

The QCAI is administered within one class period (i.e., approximately 40-45 minutes). The data analyzed in this study were collected during QCAI administrations in the Fall and Spring of the 1990-91, 1991-92, and 1992-93, and the Spring of the 1993-94 instructional years. Students who were tested on more than one occasion during the three years received a different form of the QCAI on each administration[5].

Scoring Student Responses

A focused holistic scoring method was used for scoring the student responses to each task. This was accomplished by first developing a general scoring rubric that reflected the conceptual framework used for constructing the assessment tasks (Lane, 1993). The general scoring rubric incorporates three interrelated components: mathematical conceptual and procedural knowledge, strategic knowledge, and communication. In

---

[4] A discussion of the design and administration considerations of the QCAI Bilingual Version is presented in Lane and Silver (1995).

[5] The forms were randomly distributed within each class in the schools participating in QUASAR in the fall of 1990, and thereafter each student received a different form on each administration occasion (Lane, Stone, Ankenmann, & Liu, 1994). The use of this sampling approach allows for the assessment of students in a relatively short time frame, thereby keeping interruptions to the instructional process minimal; minimizes the occurrence of practice effects; avoids the problems associated with sampling only a small number of tasks (e.g., Mehrens, 1992); and affords valid generalizations about students' mathematical proficiency at the school level.

developing the general scoring rubric, criteria representing the three interrelated components were specified for each of five score levels (0-4). Five score levels were used to facilitate capturing various levels of student understanding. The general scoring rubric is provided in Figure 2.

Insert Figure 2 about here

Based on the specified criteria at each score level a specific rubric was developed for each task. The emphasis on each component for a specific rubric is dependent on the cognitive demands on the task. The criteria specified at each score level for each specific rubric is guided by theoretical views on the acquisition of mathematical knowledge and processes assessed by the task, and the examination of actual student responses to the task. This scoring procedure allows the assessment of differential levels of students' mathematical proficiency.

Student responses were rated by middle school mathematics teachers. The raters scored the student responses after they were formally trained. First, the general rubric was presented and discussed. Then a specific rubric and pre-scored student responses were presented and discussed. The raters then practiced scoring student responses, and their scores were discussed in relation to the scores previously assigned by the assessment team. Finally, the raters scored the actual student responses. Each response was scored independently by two raters. If the raters disagreed by more than one point, the assessment team rated the student response and it was this score that was used in subsequent analyses.

<u>Generalizability Analyses</u>

In examining the generalizability (i. e., reliability) of the scores, both intertask and interrater consistency were addressed. Lane, Liu, Ankenmann, and Stone (in press) reported generalizability coefficients for a person-by-task design, using 9 tasks, that ranged from .69 to .84 depending upon the form and grade level. Consistent with previous studies examining intertask consistency for performance assessments in other subject matters (e.g., Shavelson, Baxter, and Pine (1991)), the variability due to the person-by-task interaction accounted for between 53% and 73% of the total variability. Lane et al. also found that the variances due to rater, rater x person, and rater x task were negligible, indicating that the raters were consistent in applying the criteria specified in the rubrics for assigning scores to the student responses.

Data Analyses

Student scores from the set of 11 QCAI tasks that appeared on both the 6th/7th and 8th grade versions of the QCAI and that were administered in each of the first four years of the project were used for the analyses. The analyses compared changes in performance on the QCAI for subgroups of students within each of the two sites. It should be noted that the sample of students is not necessarily the same students across the instructional years because of new students entering the schools, students leaving the schools, and student absences on the days the QCAI was administered. It was not possible to use only those students who were in the program for all of the years because of small sample sizes for some of the subgroups.

In these analyses, the average percentage of student responses that obtained the two most proficient score levels (3 or 4) was examined over time. Student responses that obtained a level of 4 based on the focused holistic scoring procedure demonstrated correct and complete mathematical understanding of the problem, the use of appropriate solution processes and representations, and coherent, complete mathematical reasoning. Student responses that are not of sufficiently high quality to receive a score level of 4, but contain only a fairly minor error in mathematical knowledge, solution strategy, or reasoning processes would receive a score of 3.

In general, the data analyzed were obtained during the administration of the QCAI in the Fall and Spring of the 1990-1991, 1991-1992, 1992-1993, and the Spring of the 1993-1994 instructional years. To examine changes in student performance across the first 3 years of the project, one set of analyses begins with sixth grade student responses in the 1990-91 school year, seventh-grade student responses in the 1991-92 school year, and eighth-grade student responses in the 1992-1993 school year. To examine changes in student performance across the second 3 years of the project, another set of analyses begins with sixth-grade student responses in the 1991-92 school year, seventh-grade student responses in the 1992-1993 school year, and eighth-grade student responses in the 1993-1994 school year.

RESULTS

Performance of Caucasian and African American Students in School A

Two analyses of changes in student performance for African American and Caucasian students were undertaken for School A. One analysis was done for the longitudinal cohort of students tested in 6th grade in Fall 1990, in 6th grade in Spring 1991, in 7th grade in Spring 1992, and in 8th grade in Spring 1993.[6] In this analysis, the average number of Caucasian students responding to each task was 24, 25, 24, and 22 and the average number of African American students was 9, 8, 8, and 8 in the Fall 1990, Spring 1991, Spring 1992, and Spring 1993, respectively. It should be noted that not

[6] Due to high rates of student turnover, which are typical for schools located in poor urban communities, the cohort group described here and others described throughout this paper are not "true" longitudinal cohorts. In fact, of the students tested in the spring of grade 8 in these cohorts, approximately half were not in the sample in the fall of grade 6. Because the numbers of students in the sample at each point in time is fairly small, however, it is not practical to examine only those students who remain in the cohort over time. The inclusion of all tested students in the analyses reported here undoubtedly depresses the overall achievement levels, since Lane and Silver (1994) analyzed the performance of a "true" longitudinal cohort at one QUASAR site and found that this group achieved greater gains than the overall gains reported for the entire sample of stdents tested at the school. Although the overall performance levels for the entire group of students is likely lower than that for the true cohort, there is no resason to assume that this will affect relative performance differences for various subgroups.

each student responded to each task because each student received only one of the four forms and the 11 tasks were distributed across the four forms. Thus, the total number of students used in each analysis is about four times the average number of students responding to each task.

The second analysis was done for the 6th grade students in the Fall 1991, 6th grade students in the Spring 1992, 7th grade students in the Spring 1993, and 8th grade students in the Spring 1994. In this analyses, the average number of Caucasian students responding to each task was 24, 25, 25, and 19 and the average number of African American students was 9, 9, 7, and 7 in the Fall 1991, Spring 1992, Spring 1993, and Spring 1994, respectively. These analyses allow for the examination of the impact of the instructional program on Caucasian and African American student performance in School A for the first three years of the project as compared to the second three years of the project. An increase in student performance during the second three years of the project would provide some evidence regarding positive changes in the implementation of the innovative instructional programs during the project years.

For School A, Figure 3 shows the percentage of Caucasian and African American student responses that were scored at the two highest score levels across the 1990-1993 time period. At each administration occasion, Caucasian students performed better than African American students. The percentage point differences between the 6th grade student performances in Fall 1990 and 8th grade student performances in Spring 1993, however, were 36 for Caucasian students and 40 for African American students. Although Caucasian 6th grade students performed better than African American 6th grade students in the fall of the first year of the project, the Caucasian and African American students had similar gain scores across the first three years of the project. In fact, the gain score for African American students is slightly higher than the gain score for Caucasian students.

The percentage of African American and Caucasian student responses that were scored at the two most proficient score levels across the 1991-1994 time period for School A is also shown in Figure 1. The difference between the 6th grade Caucasian student performances in Fall 1991 and 8th grade Caucasian student performances in Spring 1994 is 43 percentage points. This gain score is 7 percentage points higher than the gain score for Caucasian students in the 1990-1993 analysis. Thus, this finding may indicate that impact of the instructional program on Caucasian student proficiency was greater during the second 3 years than the first 3 years.

For African American students, however, the percentage point difference between their performance in 6th grade in Fall 1991 and 8th grade in Spring 1994 is 33. This gain score is 10 percentage points lower than the Caucasian students' gain score during the same time period. In addition, the African American students' gain score of 33 for the 1991-94 time period is 7 percentage points lower than their gain score of 40 for the 1990-93 time period. This low gain score during the 1991-94 time period for the African American students is due to the lack of progress for 8th grade students in the 1993-94 school year. The gain score for 8th grade African American students in the 1993-94 year was only 1 percentage point. When the 8th grade data are excluded from the analyses the percentage point differences between the 6th grade student performances in Fall 1991 and 7th grade student performances in Spring 1993 are 31 for Caucasian students and 32 for African American students. Thus, the performance gains are similar for African American and Caucasian 6th and 7th grade students during the 1991-93 time period.

The lack of gain for 8th grade African American students during the 1993-94 year appears to be due to differential opportunity to learn from the curriculum and instruction across the classes in School A. In order to afford more students an opportunity to take a course in grade 8 with a clear focus on algebra, the 8th grade curriculum at School A was more differentiated in 1993-94 than in the previous year. Four 8th grade classes were designated as algebra classes, and they studied a conceptually rich curriculum intended to develop algebraic thinking and reasoning. An unfortunate consequence of this decision to create four sections of algebra at grade 8 was the unintended recurrence of a form of instructional "tracking", which the teachers and principal had been successful in eliminating in the school several years earlier. This "tracking" resulted from the fact that the non-algebra classes tended to contain large numbers of students who had been less successful with challenging mathematical tasks in the past. Thus, the teachers of these classes found that the classes were generally less able to handle the more cognitively challenging algebra tasks than was true in the more heterogeneous classes in the previous year (Smith, personal communication, October 1994). An examination of the ethnic composition of the classes receiving these two different curricula provides a likely explanation for the poorer performance gain for 8th grade African American students, since 68% of the Caucasian 8th grade students were in the classes that were more challenging and only 30% of the African American students were in those classes. Thus, the poorer performance gains made by 8th grade African American students is likely explained by the fact that the majority of 8th grade African American students did not have the same educational opportunity as the majority of the Caucasian 8th grade students.

Since performance gains were similar for Caucasian and African American students during the 1992-93 year when the curricular and instructional focus was far less differentiated in 8th grade, it is less likely that the differences noted for the 1993-94 school year are due to differential capacity to benefit from instruction oriented toward high-level cognitive objectives. Since teachers at the school commented at the end of that school year that they were disappointed in the differentiated instruction that occurred in the two sets of eighth-grade classes, they decided to group students heterogeneously and blend algebraic material into the curriculum for all 8th grade classes in the 1994-95 school year. If our explanation of the reason for the difference noted for the student subgroups in grade 8 in 1993-94 is correct, we would expect to see similar gains for Caucasian and African American students during the 1994-95 school year.

Performance for Students in English-speaking Classes and Bilingual Classes in School B

Differential changes in performance for students in English-speaking classes and students in bilingual classes were examined in School B. The teacher for the bilingual classes instructed in both Spanish and English and students in the bilingual classes used both Spanish and English to communicate their mathematical thinking and reasoning. Similar to School A, two analyses were undertaken. One analysis was done for 6th grade students in the Fall 1990, 6th grade students in the Spring 1991, 7th grade students in the Spring of 1992, and 8th grade students in the Spring of 1993. In this analysis, the average number of students in English-speaking classes who responded to each task was 37, 37, 48, and 46 and the average number of students in bilingual classes who responded to each task was 5. 7 8, and 7 in the Fall 1990, Spring 1991, Spring 1992, and Spring 1993, respectively. The second analysis was done for the 6th grade students in the Fall 1991 and Spring 1992, 7th grade students in the Spring 1993, and 8th grade students in the Spring 1994. In this analysis, the average number of students in English-speaking classes who responded to each task was 28, 28, 55, and 51 and the average number of students in bilingual classes who responded to each task was 5, 6, 8, and 6 in the Fall 1990, Spring 1991, Spring 1992, and Spring 1993, respectively. It should be noted that over 80% of the Latino students were in the bilingual classes.

Figure 4 shows the average percentage of student responses that were scored at the two most proficient score levels across the 1990-1993 time period for the English-speaking and bilingual classes. The percentage point differences between the average 6th grade student performances in Fall 1990 and average 8th grade student performances in Spring 1993 were 26 for English-speaking classes and 31 for bilingual classes. Although 6th grade students in bilingual classes were, on average, 6 percentage points below students in the 6th grade English-speaking classes in the Fall 1990, the average for 8th grade students in bilingual classes was only 1 perc entage point below the average for 8th grade students in English-speaking classes in the Sp ing 1993.

The average percentage of student responses in the English-speaking and bilingual classes that were scored at the two most proficient score levels across the 1991-1994 time period for School B is also shown in Figure 4. The difference between the 6th grade student performance in Fall 1: 1 and 8th grade student performance in Spring 1994 is 31 percentage points for English-speaking classes and 36 percentage points for bilingual classes. Thus, although the 6th grade students in the bilingual classes were 10 percentage points behind the 6th grade students in the English-speaking classes in the Fall 1990, the 8th grade students in the bilingual classes were only 5 percentage points behind the 8th grade students in the English-speaking classes in the Spring 1994.

It should be noted that the average performance gains for students in both the English-speaking and bilingual classes were 5 percentage points higher in the 1991-1994 time period as compared to the 1990-1993 time period. The greater gains for the second three years of the project may be a result of the instructional programs being more established and implemented more effectively by the teachers. In addition, the percentage point gains for the students in the bilingual classes were 5 percentage points higher than the students in the English-speaking classes for both the 1990-93 and 1991-94 time period. Further, there was one bilingual teacher for the 6th, 7th, and 8th grade classes from Fall 1990 to Spring 1993. This bilingual teacher, however, was not at the school during the 1993-1994 instructional year and was replaced by a new teacher, and in this year the least gain was made by students in the bilingual classes.

Performance of Caucasian and African American Students in School B

Because nearly all students in the English-speaking classes were either Caucasian or African American, an analysis was undertaken to examine whether there were differences in gain scores for these two subgroups. In the analysis for the 1990-1993 cohort, the average number of Caucasian students responding to each task was 9, 8, 7, and 8 and the average number of African American students was 24, 23, 31, and 30, in the Fall 1990, Spring 1991, Spring 1992, and Spring 1993, respectively. In the analysis for the 1991-1994 cohrt, the average number of Caucasian students responding to each task was 5, 6, 9, and 8 and the average number of African American students was 18, 18, 38, and 35 in the Fall 1991, Spring 1992, Spring 1993, and Spring 1994, respectively.

Figure 5 shows the percentage of African American and Caucasian student responses that were scored at the two most proficient score levels across the 1990-1993 and 1991-1994 time periods, respectively. At each administration occasion, Caucasian students performed better than the African American students. The percentage point differences between the 6th grade student performances in Fall 1990 and 8th grade student performances in Spring 1993 were 32 for Caucasian students and 27 for African American students, whereas the difference between the 6th grade student performance in Fall 1991 and 8th grade student performance in Spring 1994 is 31 percentage points for Caucasian students and 30 percentage points for African American students. Thus, although the African American 6th grade students performed substantially lower than the Caucasian students in the fall, the two groups had similar gain scores across each of the sets of three years.

SUMMARY

In the context of a mathematics instructional reform project such as QUASAR, it is imperative to examine whether the educational opportunities provided to various ethnic and linguistic subgroups of students are similar. The results of this study indicate that the instructional programs at Schools A and B provide educational opportunities that promote similar gains in mathematical thinking and reasoning for both African American and Caucasian students. The performance of three of the four longitudinal cohorts of students examined here showed a clear pattern of parallel gains by African American and

Caucasian students in the two schools. The lone exception was for the School A cohort that included the 8th grade students who received differential instruction in the 1993-1994 school year. But, even in this cohort, the gains made for the first two years, when instrctional opportunity was equitable, the gains for the two subgroups were quite similar. Moreover, it is important to note that, although a performance gap exists for the two subgroups, the gap did not widen over grades 6-8, as might have been expected from previous research. Thus, the results suggest that these QUASAR schools are providing students with equitable access to educational opportunities and that this access has led to similar proficiency gains in mathematical thinking and reasoning for African American and Caucasian students.

The results from the present study also indicate that high quality mathematics instruction can be made available to students in bilingual classes as well as in monolingual classes. Although students in the bilingual classes in School B had lower initial proficiency levels than students in the English-speaking classes, their gains in proficiency exceeded the gains made by the students in the English-speaking classes. Thus, the gap between the mathematics performance for 8th grade students in English-speaking and bilingual classes was narrower than the gap between the mathematics performance for 6th grade students in English-speaking and bilingual classes. These results indicate that the mathematics instructional program, which is oreiented toward high-level cognitive objectives, is as effective for bilingual students as it is for English-speaking students at School B. Thus, given similar educational opportunities to think and reason about mathematical ideas, bilingual students can benefit as much as English-speaking students. In summary, it appears as if the instructional programs at these two QUASAR schools are promoting enhanced mathematical thinking and reasoning for African American and Latino students as well as for Caucasian students.

The similar proficiency gains for these subgroups of students also lends support for the validity of the QCAI. Evidence is needed to ensure that assessments, regardless of item type, are equally valid for all subgroups being assessed. The results of this study indicate that, despite differences in initial proficiency levels for various ethnic and linguistic subgroups of students, if the same educational opportunities are provided, the subgroups can attain similar gains in proficiency. In fact, in some cases the gains in proficiency were greater for some African American and bilingual subgroups of students than Caucasian and English-speaking subgroups, respectively. Thus, the findings provide evidence that the assessment is sensitive to measuring changes in student performance for various ethnic and linguistic subgroups.

REFERENCES

Baker, E. L., O'Neil, H. F., & Linn, R. L. (1991). Policy and validity prospects for performance-based assessment. Paper presented at the annual meeting of the American Psychological Association.

Barr, R., & Dreeben, R. (1983). How schools work. Chicago: University of Chicago Press.

College Entrance Examination Board (1985). Equality and excellence: The educational status of Black Americans. New York: Author.

Darling-Hammond, L. (1995). Equity issues in performance-based assessment. In M. T. Nettles & A. L. Nettles, Equity and excellence in educational testing and assessment (pp. 89-114). Boston, MA: Kluwer.

Darling-Hammond, L., & Snyder, J. (1991). Traditions of curriculum inquiry: The scientific tradition. In P. Jackson (Ed.), Handbook of research on curriculum. New York: Macmillan.

De Avila, E. A. (1988). Bilingualism, cognitive function, and language minority group membership. In R. R. Cocking & J. P. Mestre (Eds.), Linguistic and cultural influences on learning mathematics (pp. 101-121). Hillsdale, NJ: Lawrence Erlbaum.

Dunbar, S. B. (1987, April). Comparability of indirect measures of writing as predictors of writing performance across demographic groups. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Dunbar, S. B.., Koretz, & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education. 4, 289-304.

Feinberg, L. (1990, Fall). Multiple choice and its critics. The College Board Review, No. 157.

Gross, S. (1988, July). Participation and performance of women and minorities in mathematics (Report of National Science Foundation project no. MDR-8470384). Rockville, MD: Montgomery,County Public Schools.

Jones, L. V. (1984). White-black achievement differences: The narrowing gap. American Psychologist, 39, 1207-1213.

Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment. Educational Measurement: Issues and Practice, 12(2), 16-23.

Lane, S., Liu, M., Ankenmann, R. D., & Stone, C. A. (in press). Generalizability and validity of a mathematics performance assessment. Journal of Educational Measurement.

Lane, S., Parke, C., & Moskal, B. (1992). Principles for developing performance assessments. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lane, S., & Silver, E. A. (1995). Equity and validity considerations in the design and implementation of a mathematics performance assessment: The experience of the QUASAR project. In M. T. Nettles & A. L. Nettles, Equity and excellence in educational testing and assessment (pp. 185-220). Boston, MA: Kluwer.

Lane, S., & Silver, E. A. (1994, April). Examining students' capacities for mathematical thinking and reasoning in the QUASAR project. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Lane, S., Stone, C. A., Ankenmann, R. D., & Liu, M. (1994). Reliability and validity of a mathematics performance assessment. International Journal of Educational Research, 21(3), 247-262.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21.

Magone, M. E., Cai, J., Silver, E. A., & Wang, N. (1994). Validity evidence for cognitive complexity of performance assessments: An analysis of selected QUASAR tasks. International Journal of Educational Research, 21(3), 317-340.

Magone, M. E., Wang, N., Cai, J., & Lane, S. (1993). An analysis of the cognitive complexity of QUASAR's performance assessment tasks and their sensitivity to measuring changes in students thinking. A paper presented in the symposium "Assessing performance assessments: Do they withstand empirical scrutiny?" at the 1993 Annual Meeting of the American Educational Research Association, Atlanta, GA.

Mehrens, W. A. (1992). Using performance assessment for accountability purposes. Educational Measurement: Issues and Practice, 11(1), 3-9, 20.

Mullis, I. V. S., Dossey, J. A., Cambell, J. R., Gentile, C. A., O'Sullivan, C., & Latham, A. S. (1994). NAEP 1992 trends in academic progress. Washington, DC: U.S. Deaprtment of Education, Office of Educational Research and Improvement.

National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics (1991). Professional standards for teaching mathematics. Reston, VA: Author.

National Research Council (1989). Everybody counts. Washington, DC: National Academy of Sciences.

Neill, M. (1995). Some prerequisities for the establishment of equitable, inclusive multicultural assessment systems. In M. T. Nettles & A. L. Netties, Equity and excellence in educational testing and assessment (pp. 115-157). Boston, MA: Kluwer.

Oakes, J. (1990). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science. Santa Monica: RAND Corporation.

O'Connor, M. C. (1989). Aspects of differential performance by minorities on standardized tests: Linguistic and sociocultural factors. In B. R. Gifford (Ed.), Test policy and test performance: Education, language, and culture (pp. 129-181). Boston, MA: Kluwer.

O'Connor, M. C. (1992). (footnote on pg. 4 of paper)

Resnick, L. B., & Resnick, D. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), Future assessments: Changing views of aptitude, achievement and instruction. Boston, MA: Kluwer.

Secada, W. G. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. A. Grouws, Handbook of research on mathematics teaching and learning. New York, Macmillan.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. Applied Measurement in Education, 4(4), 347-362.

Silver, E. A. (1992). Assessment and mathematics education reform in the United States. International Journal of Educational Research, 17(5), 489-502.

Silver, E. A. (1994). Mathematical thinking and reasoning for all students: Moving from rhetoric to reality. In D. F. Robitaille, D. H. Wheeler, & C. Kieran (Eds.), Selected lectures for the 7th International Congress on Mathematical Education (pp. 311-326). Sainte-Foy, Quebec: Les Presses De L'Université Laval.

Silver, E. A. (1993, March). QUASAR Project Summary. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Silver, E. A., & Lane, S. (1993). Assessment in the context of mathematics instruction reform: The design of assessment in the QUASAR project. In M. Niss (Ed.), Assessment in mathematics education and its effects (pp. 59-70). London: Kluwer Academic.

Stein, M. K., Grover, B. W., & Henningsen, M. (In press). Enhanced instruction as a means of building student capacity for mathematical thinking and reasoning. American Educational Research Journal.

Stodolsky, S. (1988). The subject matters: Classroom activity in mathematics and social studies. Chicago: University of Chicago Press.

Wang, N. & Lane, S. (1994). Detection of differential item functioning (DIF) in a mathematics performance assessment. Paper presented at the annual meeting of the National Council of Measurement in Education, New Orleans, LA.

Winfield, L. F. (1995). Performance-based assessments: Contributor or detractor to equity?. In M. T. Nettles & A. L. Nettles, Equity and excellence in educational testing and assessment (pp. 221-241). Boston, MA: Kluwer.

Wong Fillmore, L., & Valadez, C. (1986). Teaching bilingual learners. In M. C. Wittrock (Ed.), Handbook of research on teaching (pp. 648-685). New York: Macmillan.

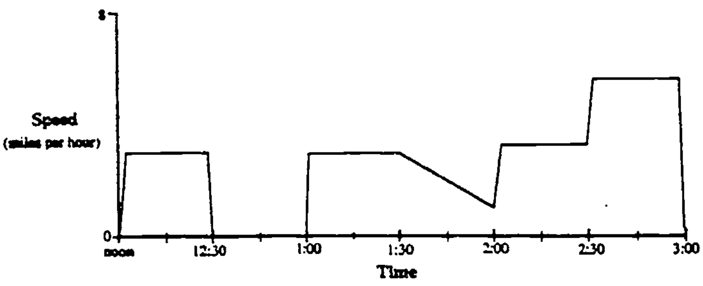
Figure 1

Four OCAI Release Tasks

OCAI Graph Interpretation Task

Use the following information and the graph to write a story about Tony's walk.

At noon, Tony started walking to his grandmother's house. He arrived at her house at 3:00. The graph below shows Tony's speed in miles per hour throughout his walk.



Write a story about Tony's walk. In your story, describe what Tony might have been doing at the different times.

OCAI Decimal Value Task

Circle the number that has the greatest value.

.08 .8 .080 00800

Explain your answer.

OCAI Number Theory Task

Yolanda was telling her brother Damian about what she did in math class.

Yolanda said, "Damian, I used blocks in my math class today. When I grouped the blocks in groups of 2, I had 1 block left over. When I grouped the blocks in groups of 3, I had 1 block left over. And when I grouped the blocks in groups of 4, I still had 1 block left over."

Damian asked, "How many blocks did you have?"

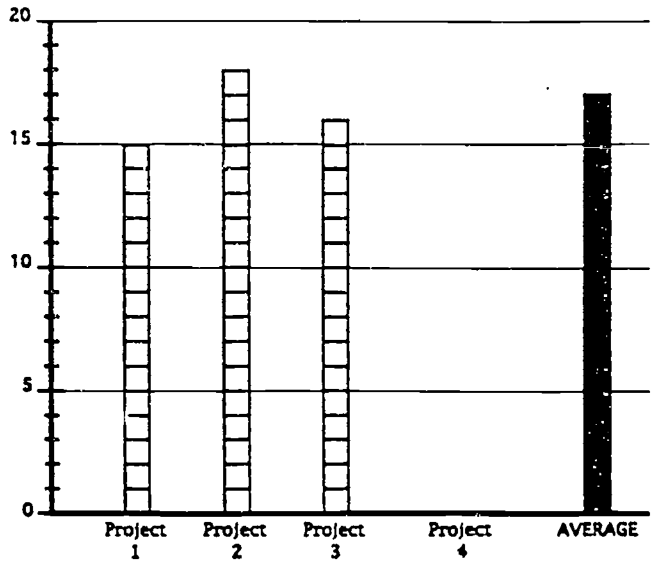
What was Yolanda's answer to her brother's question?

Show your work.

Answer: ______________

OCAI Average Score Task

Anita has four 20-point projects for science class. Anita's scores on the first 3 projects are shown below.



A. What score must Anita get on Project 4 so that her average for the four projects is 17?

Answer: ______________ You may draw your answer on the graph.

B. Explain how you found your answer.

Figure 2

General Scoring Rubric

Score Level 4

Mathematical knowledge
- Shows understanding of the problem's mathematical concepts and principles;
- Uses appropriate mathematical terminology and notations;
- Executes algorithms completely and correctly.

Strategic knowledge
- May use relevant outside information of a formal or informal nature,
- Identifies all the important elements of the problem and shows understanding of the relationships among them;
- Reflects an appropriate and systematic strategy for solving the problem;
- Gives clear evidence of a solution process, and solution process is complete and systematic.

Communication
- Gives a complete response with a clear, unambiguous explanation and/or description;
- May include an appropriate and complete diagram;
- Communicates effectively to the identified audience;
- Presents strong supporting arguments which are logically sound and complete;
- May include examples and counter-examples.

Score Level 3

Mathematical knowledge
- Shows nearly complete understanding of the problem's mathematical concepts and principles;
- Uses nearly correct mathematical terminology and notations;
- Executes algorithms completely; computations are generally correct but may contain minor errors.

Strategic knowledge
- May use relevant outside information of a formal or informal nature;
- Identifies the most important elements of the problem and shows general understanding of the relationships among them;
- Gives clear evidence of a solution process; solution process is complete or nearly complete, and systematic.

Communication
- Gives a fairly complete response with reasonably clear explanations or descriptions;
- May include a nearly complete, appropriate diagram;
- Generally communicates effectively to the identified audience;
- Presents supporting arguments which are logically sound but may contain some minor gaps.

Score Level 2

Mathematical knowledge
- Shows some understanding of the problem's mathematical concepts and principles;
- May contain serious computational errors.

Strategic knowledge
- Identifies some important elements of the problem but shows only limited understanding of the relationships among them;
- Gives some evidence of a solution process. but solution process may be incomplete or somewhat unsystematic.

Communication
- Makes significant progress towards completion of the problem, but the explanation or description may be somewhat ambiguous or unclear;
- May include a diagram which is flawed or unclear;
- Communication may be somewhat vague or difficult to interpret;
- Arguments may be incomplete or may be based on a logically unsound premise.

Score Level 1

Mathematical knowledge
- Shows very limited understanding of the problem's mathematical concepts and principles;
- May misuse or fail to use mathematical terms;
- May contain make major computational errors.

Strategic knowledge
- May attempt to use irrelevant outside information;
- Fails to identify important elements or places too much emphasis on unimportant elements;
- May reflect an inappropriate strategy for solving the problem;
- Gives incomplete evidence of a solution process; solution process may be missing, difficult to identify, or completely unsystematic.

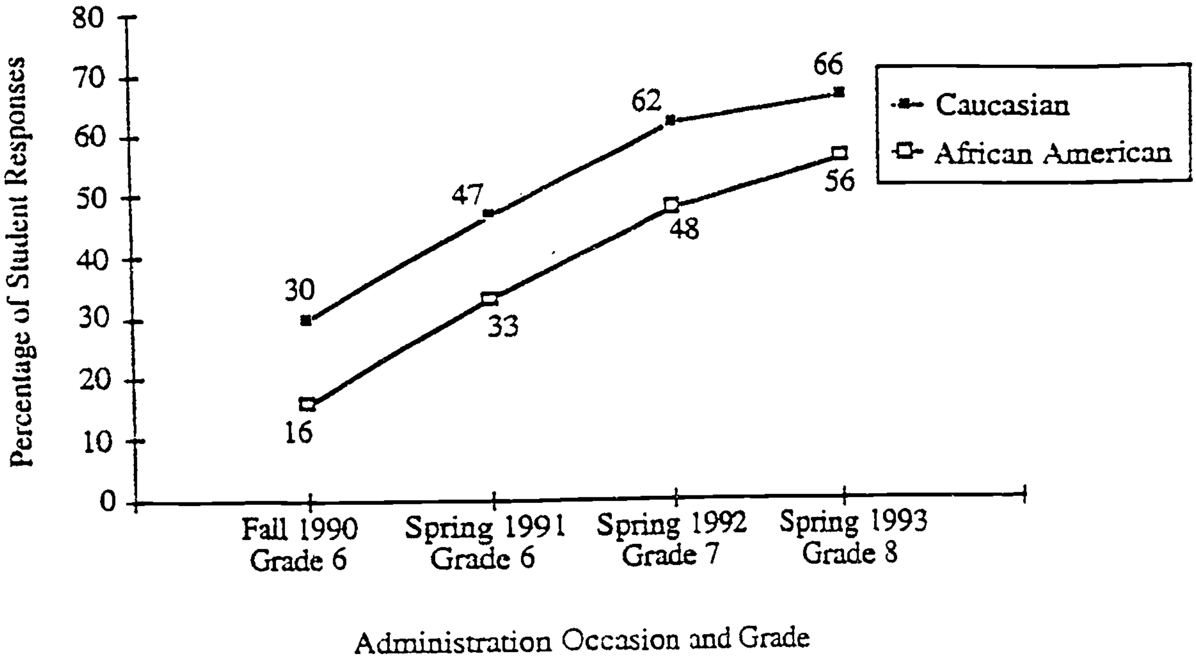
Communication
- Includes some satisfactory elements but may fail to complete or may omit significant parts of the problem;
- Explanation or description may be missing or difficult to follow;
- May include a diagram which incorrectly represents the problem situation, or diagram may be unclear and difficult to interpret.

Score Level 0

Mathematical knowledge
- Shows no understanding of the problem's mathematical concepts and principles.

Strategic knowledge
- May attempt to use irrelevant outside information;
- Fails to indicate which elements of the problem are appropriate;
- Copies part of the problem, but without attempting a solution.

Communication
- Communicates ineffectively; words do not reflect the problem;
- May include a diagram which completely misrepresents the problem situation.



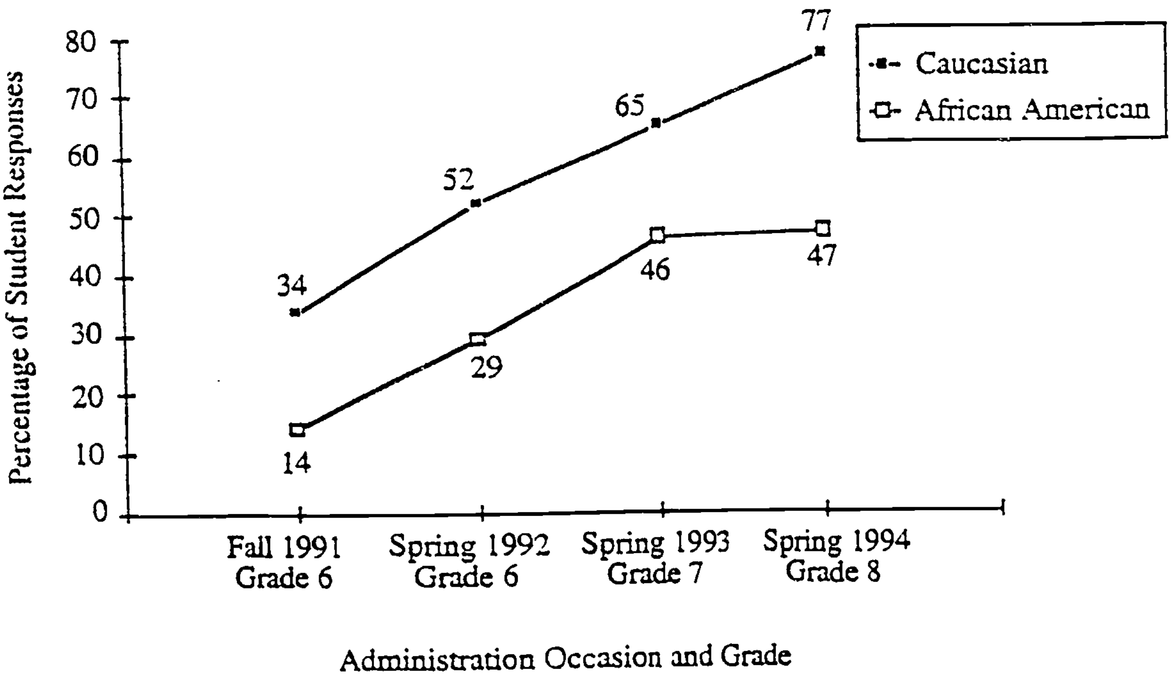
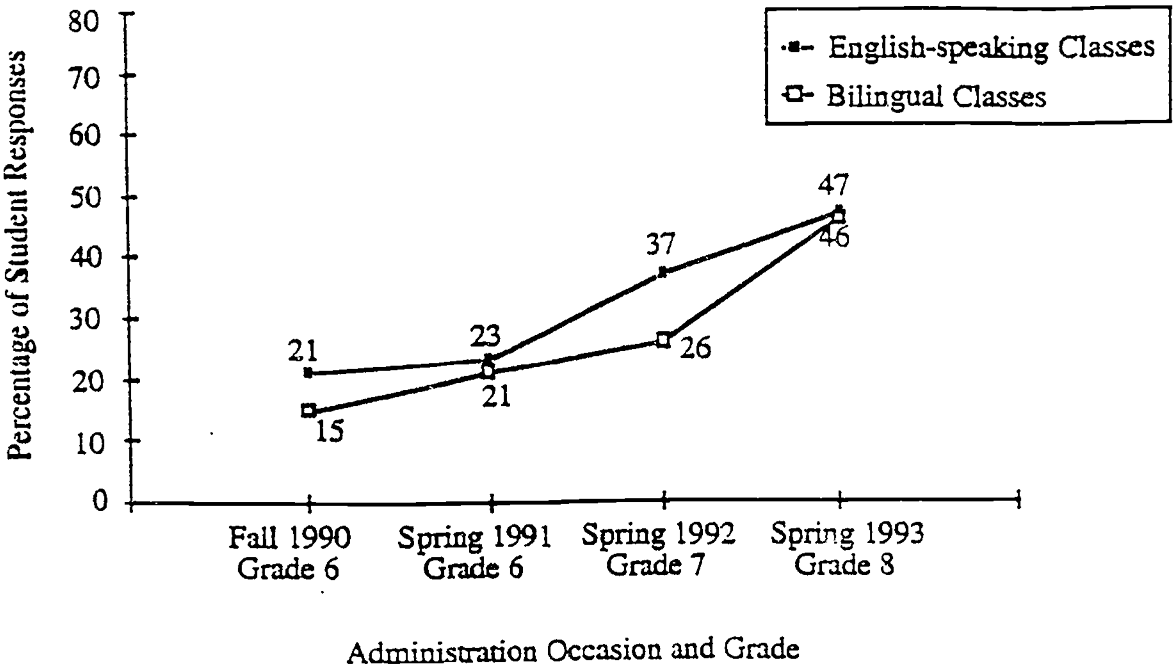


Figure 3. Average Percentage of Caucasian and African American Student Responses at the Two Most Proficient Score Levels on 11 QCAI Tasks at School A



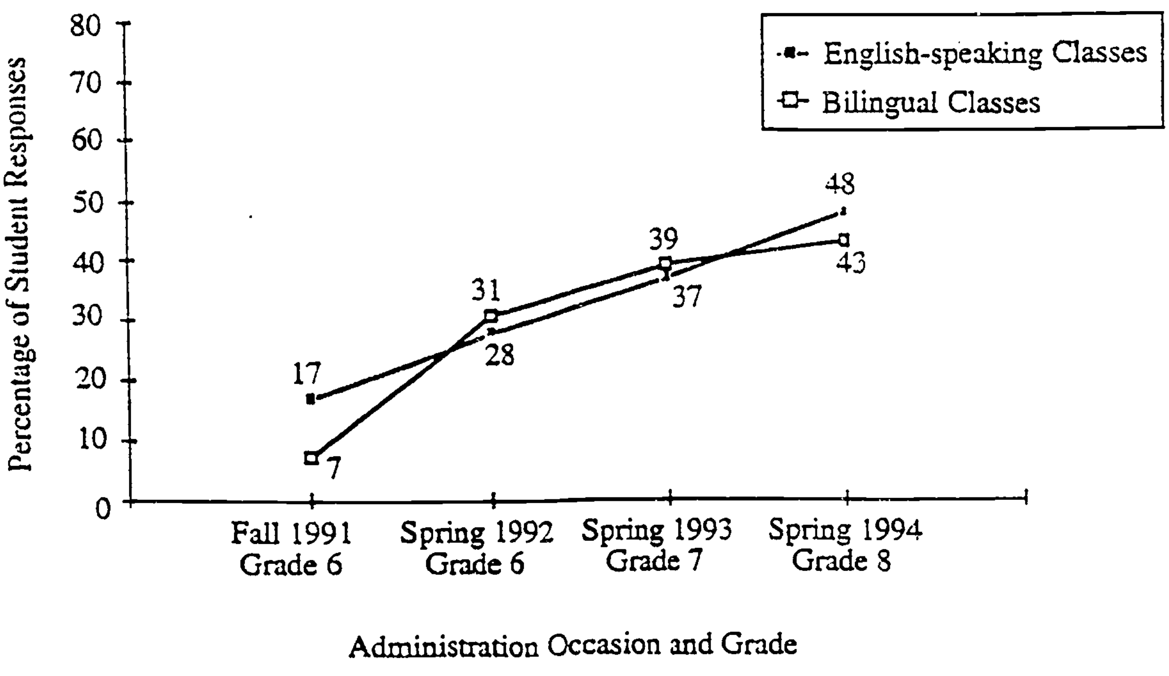
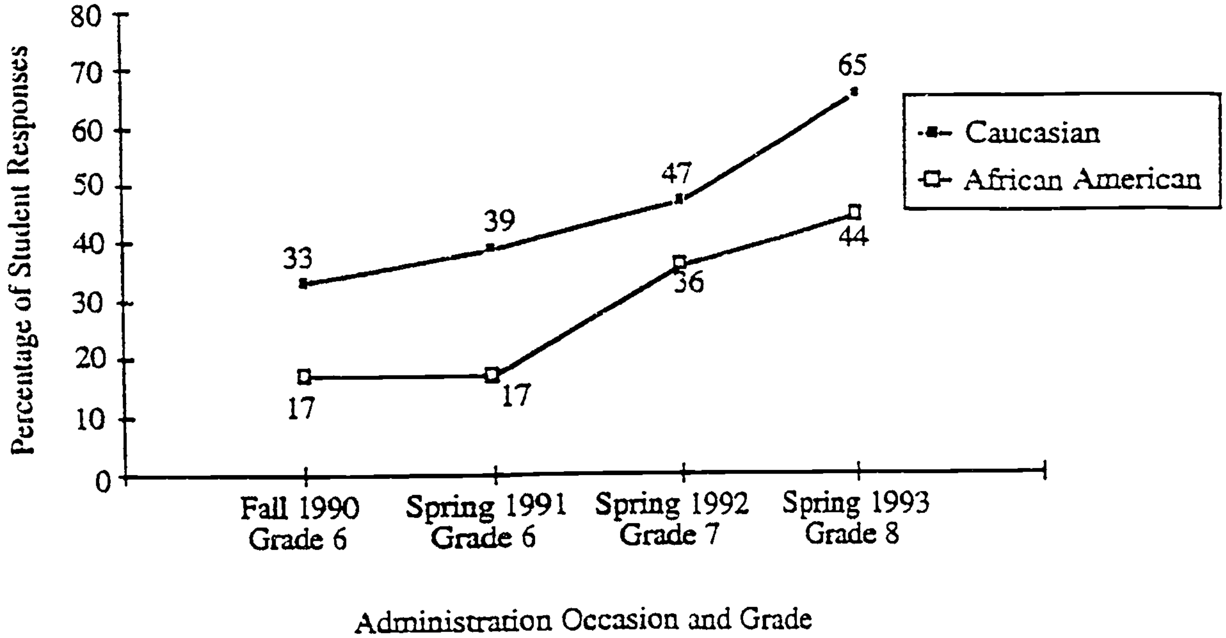


Figure 4. Average Percentage of Student Responses at the Two Most Proficient Score Levels on 11 QCAI Tasks for English-speaking Classes and Bilingual Classes at School B



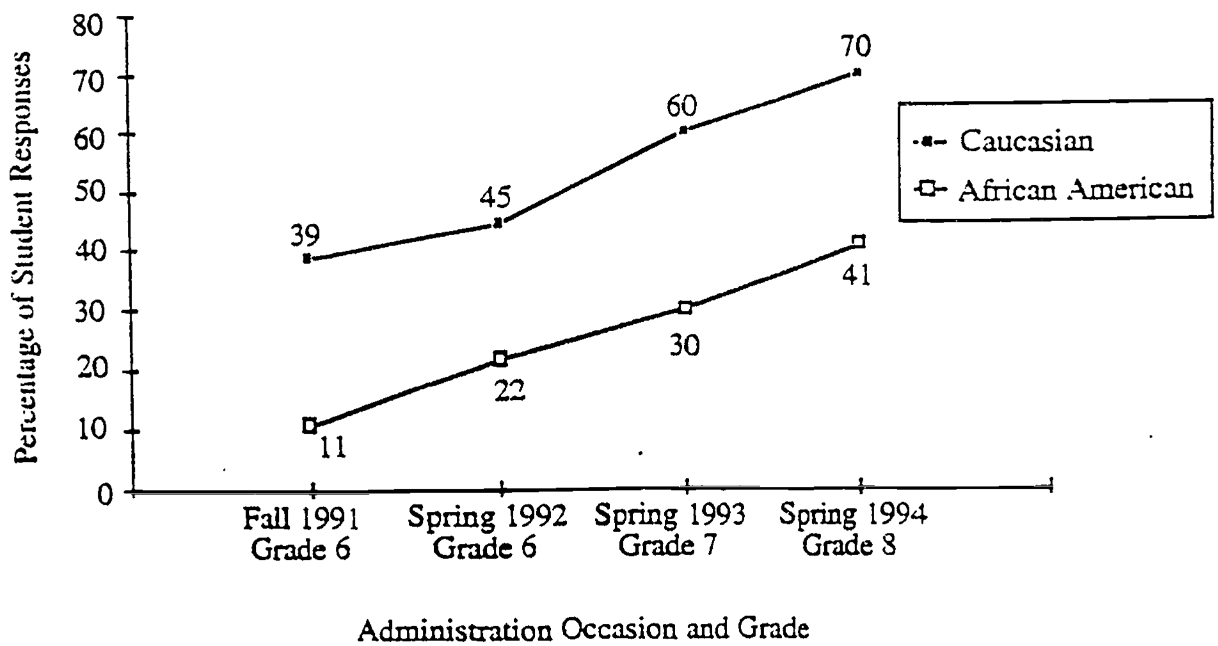


Figure 5. Average Percentage of Caucasian and African American Student Responses at the Two Most Proficient Score Levels on 11 QCAI Tasks at School B