Assessing Equity in Alternative Assessment:
An Illustration of Opportunity-to-Learn Issues

CSE Technical Report 440

Joan L. Herman, Davina C. D. Klein, and Sara T. Wakai
CRESST/University of California, Los Angeles

Tamela Heath, University of Maryland, College Park

December 1996

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-6511
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA Catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

ASSESSING EQUITY IN ALTERNATIVE ASSESSMENT: AN ILLUSTRATION OF OPPORTUNITY-TO-LEARN ISSUES

Joan L. Herman, Davina C. D. Klein, and Sara T. Wakai

UCLA, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

Tamela Heath, University of Maryland, College Park

Abstract

Based on the 1993 California Learning Assessment System (CLAS) Middle Grades Mathematics Performance Assessment, an innovative alternative assessment, the study explores whether all schools, regardless of the cultural, ethnic, or socioeconomic background of the students they serve, provide students equal opportunity to learn that which is assessed. Opportunity to learn was defined to include a range of variables likely to influence student performance, including access to resources, access to high-quality instructional content and processes, extra-school opportunities, and direct preparation for the CLAS. Data collection efforts included teacher interviews, student surveys, student retrospective think-aloud interviews, and classroom observations of the assessment administration. Researchers chose 13 schools across the state to represent three broad school categories: affluent suburban; low-SES urban; and remote, mixed SES rural. Findings highlight some differences between school types in various opportunity-to-learn measures and suggest directions for future research.

Introduction

Bolstered by research showing that traditional testing has encouraged teachers and students to focus on what is tested (Herman & Golan, 1991; Madaus, 1991; Shepard, 1991), but cognizant of findings suggesting that such a focus has distorted the curriculum for many students, narrowing it to basic, low-level skills (Herman & Dorr-Bremme, 1983; Herman & Golan, 1991; Kellaghan & Madaus, 1991; Shepard, 1991; Smith & Rottenberg, 1991), many in the educational community are looking toward new kinds of assessments to support educational reform. They seek assessments that embody rigorous standards for student accomplishment and whose use will foster instructional improvement.

Will these assessments support instructional reform and stimulate all students to achieve rigorous standards, as intended? Although any number of variables will intervene to influence the answer, in this article we highlight one critical link in the policy chain: that teachers and schools have the capability to and do provide all students with the opportunity to learn that which is assessed. Lacking such opportunity, interpretations of students' performance will be flawed, and the assessment will disadvantage those students who have not had equal opportunity to learn.

In this report, we use data collected in conjunction with the California Learning Assessment System to illustrate the substance and challenge of exploring this opportunity-to-learn issue, particularly in the context of concerns for equity in opportunity. We pose the question: Does current practice present a level playing field—do all schools, regardless of the cultural, socioeconomic, or community background of the students they serve, provide students with similar opportunity to learn that which is valued by new assessments? We have cast our definition of opportunity broadly to *illustrate* the range of variables likely to influence students' performance. Included in our definition were access to resources, such as qualified teachers and appropriate instructional tools; access to the types of instructional content and processes likely to help students develop the complex knowledge and skills required by new assessments; extra-school opportunities; and direct preparation and practice for these new assessments. In the sections that follow, we provide background on study methodology, a summary of our findings, and implications for policy and practice, including methodological and substantive challenges in assessing and assuring adequate opportunity to learn.

Methodology[1]

Assessment Context: California Learning Assessment System (CLAS) Middle Grades Mathematics Performance Assessment

California, at the time of the study, was known as a front-runner in curriculum and assessment reform. Planning for the state's new mathematics curriculum framework started in 1989, resulting in a published framework in 1992

[1] See Herman, Klein, Heath, and Wakai (1994) for additional detail about study methodology and findings.

(California State Department of Education, 1992). In assessment, the state had a history of using alternative assessment, having been an early adopter of direct writing assessment and having started its exploration of open-ended mathematics problems in 1989 (California State Department of Education, 1989).

The 1993 California Learning Assessment System (CLAS) Middle Grades Mathematics Performance Assessment, the focus here, had been under development for a number of years and was in its first year of full operation at the time of the study. Intended to support the state's curriculum framework, the assessment was designed to measure students' complex mathematical thinking, communication, and problem-solving skills. CLAS used a matrix sampling design and at the eighth-grade level had a total of eight forms. In this study we focused particularly on one of these forms, the common form, which was specially administered in all study classrooms on the day following the regular assessment. Like all of the forms, the common form consisted of two sections, the first containing two open-ended tasks and the second composed of eight multiple-choice items. The two open-ended tasks were designed to pose authentic, relevant problem situations for students to solve, asking students to explain their assumptions and thinking and to construct their answers using multiple modes of representation. The multiple-choice items were intended to assess mathematical thinking. While administrators were advised to give students whatever time they needed, designers expected that each test form would take about 45 minutes to complete, with students spending about 15 minutes on each of the open-ended tasks, and the remaining 15 minutes on the multiple-choice items.

School Sample

The study's original design sought to contrast schools across the state, serving diverse school communities. Because of equity concerns, the contrasts of particular interest were between schools serving relatively affluent, suburban communities and schools thought to be potentially at risk: those serving inner-city, economically disadvantaged communities and those in more geographically remote rural areas. In addition, because inner-city students were considered most at risk, inner-city schools were deliberately overrepresented. Within each school, three eighth-grade math classes, representing the range of eighth-grade classes typically taught at that school, were selected for study.

Built on and dependent on a larger state pilot study, the final sample consisted of 13 schools across the state, distributed over three broad categories: urban, rural, and suburban (see Table 1). The sample encompassed 27 teachers

Table 1
Breakdown of Schools Participating in the CLAS Study

Type of school	Number of schools	Number of classes
Urban	9	24
Rural	2	6
Suburban	2	6

(66.7% from urban schools, 14.8% from rural schools, and 18.5% from suburban schools) and over 800 students (58.4% from urban schools, 20.2% from rural schools, and 21.4% from suburban schools). (See Table 1 for a breakdown of the school sample.) The urban schools were all economically disadvantaged and reflected a range of ethnic diversity—principally Latino; mixed African American and Latino; mixed Asian American and White; mixed White, African American, and Latino. The suburban schools served predominantly White and some Asian American high-wealth communities. The rural schools were mixed in socioeconomic status and served mainly White and Latino students. It is important to note that participating schools were volunteers, interested in being involved in special pilot work for CLAS.

Instrumentation

Of the seven data sources used in the full study, results reported here draw primarily on the following.

Teacher interviews. Interviews provided information about teachers' educational background and teaching experience, particularly in mathematics; classroom pedagogical practices; teachers' familiarity with and the extent to which they prepared their students for CLAS-type items; calculator instruction and use in the classrooms; and teachers' reactions to the CLAS. In addition, during the interview, teachers were asked to provide researchers with (a) descriptions of major assignments given to students during the year and (b) samples of tests and quizzes given during the year. The teacher interview focused on the class that was observed and whose students completed student surveys for the study.

Student surveys. Students in sampled classrooms completed a survey on the day following the common form administration. The survey solicited students' views on a number of issues, including: their instructional experience with and specific preparation for the content and task types on the CLAS; access to calculators at home and at school; the amount of mathematics homework they completed; attitudes towards math in general; and their affective responses to open-ended tasks compared to multiple-choice tasks. Between the sets of student and teacher questions, there was intentional overlap in the areas of instructional practices and content coverage. Substantive and factor analyses were used to aggregate students' responses to individual items related to opportunity to learn into more stable, conceptually distinct subscales. Our analysis of collected instructional artifacts generally confirmed the validity of these subscales. For example, aggregate ratings of the communication requirements of teacher assignments and assessment correlated .69 with the communication subscale created from student responses. Similarly, material ratings of opportunities for applied problem solving correlated .52 with student reports.

Retrospective student interviews. In-depth student interviews were conducted with six students randomly chosen from each classroom. The individual student interviews allowed us to obtain more detailed information on student responses to the open-ended and multiple-choice tasks included in the assessment. Think-aloud protocols asked students to recreate their thinking processes and expectations as they approached and tried to solve one of the two open-ended tasks and the first multiple-choice item included on the common form. Students were asked to explain how they thought each task would be scored, their level of preparation for specific items, and their relative preferences, along a number of affective dimensions, for open-ended versus multiple-choice problems.

Classroom observations. In each study classroom, two trained data collectors observed in each study classroom as the common form was administered. Using a standard protocol, two observers collected information on administration conditions, students' reactions to the assessment, their engagement level, their use of calculators, and their use of time and completion of the assessment.

Data were analyzed and statistical tests run using the classroom as the unit of analysis. Analysis of variance and chi square techniques were used to explore differences in classroom experiences by school type (i.e., urban, rural, and suburban). Note that the relatively small sample of teachers and classrooms represents a significant constraint on the power of our analyses and an important caveat in interpreting results.

Results

Results are illustrated in three general areas: students' access to quality resources; students' access to instruction consonant with the CLAS objectives; and students' preparation for the CLAS.

Access to Quality Resources

With regard to access to quality resources, the study looked at both teachers' preparation for teaching a "thinking curriculum" in mathematics and students' access to adequate instructional materials. In the first area, teacher interviews provided data on teachers' undergraduate fields, mathematics teaching credential status, years of experience teaching mathematics, participation in recent professional development that would likely prepare them in the content and instructional practices of a "thinking curriculum," and preparation for the CLAS itself. Students' access to calculators and use of recent textbooks served as proxies for adequacy of instructional materials.

Teacher preparation. Although only half of our teacher sample had either majored (23.1%) or minored (26.9%) in a mathematics field (including engineering and computer science), the majority of teachers (69.2%) had credentials to teach mathematics. Rural teachers were significantly less likely to have such certification; 82% of the urban teachers and 80% of the suburban teachers were so certified, but only 25% of the rural teachers were. Similarly, suburban and urban teachers were more likely than rural teachers to have majored or minored in mathematics as undergraduates, with no rural teachers claiming an undergraduate degree in mathematics.

Similar patterns emerged when data on in-service education were examined. Overall, 65.4% of the teachers had participated recently in more than 35 hours of in-service education in mathematics and mathematics education. Urban and suburban teachers were more likely to have done so: 71% of the urban teachers and 80% of the suburban teachers reported spending more than 35 hours over the last three years in in-service education in mathematics or its teaching, while only 25% of the rural teachers reported that level of activity. No differences were found across school type in years of teacher experience teaching mathematics, with a mean of 11 years for the total sample.

With regard to teachers' specific preparation for the assessment, the majority of teachers reported participating in two to three extended workshops and other special sessions acquainting them with the CLAS and the types of mathematical thinking, communication, and problem solving contained on the assessment, in addition to reading advanced written materials, samplers, and directions for administration. Yet despite this orientation, teachers in general did not report being highly confident about their preparation to teach CLAS-type objectives, with no more than half the teachers representing each school type expressing that they felt "very well" prepared (see Table 2).

Table 2

Percentage of Teachers Expressing Different Levels of Preparation to Teach the CLAS by School Type (Teacher Interview Results)

Type of school	Not well	OK	Very well
Urban	8.7	60.9	30.4
Rural	16.7	33.3	50.0
Suburban	16.7	33.3	50.0
Totals	11.4	51.4	37.1

Instructional resources. Because the National Council of Teachers of Mathematics (NCTM) standards and the California Curriculum Framework in Mathematics are relatively new, it is unlikely that older texts are well aligned with the reform ideas of these new standards. We thus viewed recency of texts as an indicator of access to relevant instruction resources; without access to recent texts, teachers lack an important supplement to their own mathematical background and to their thinking about effective instructional activities. The data indicate that students in urban classrooms were less likely to have recent texts than those in other schools in our sample; $F(2, 27) = 5.30$, $p = .01$.

Access to calculators. We used access to calculators as another indicator of the availability of instructional tools that reflect NCTM standards and the California framework. Observation data suggested no major differences between schools in students' access to basic calculators, although, according to teacher reports, urban schools were more likely to provide them for students than were rural and suburban schools, and students in suburban schools were more likely to bring them from home. Of note is that for all types of schools, more than 90% of the students have calculators at home. The difference, however, is in the type of calculators students have available to them at home: 62.7% of the suburban students have scientific calculators (as opposed to simple calculators) at home, whereas only 43.5% of the urban students and 31.5% of the rural students have such calculators at home; $\chi^2(2, N = 632) = 24.83$, $p < .001$. Although scientific calculators were not required for the CLAS, the availability of sophisticated calculators may indicate more familiarity and ease of use with such tools.

Access to Learning Opportunities Appropriate to CLAS-Type Objectives

Perceived fit between instructional practices and the CLAS. Asked how well their instruction aligned with the material on the CLAS assessment, two-thirds of the suburban teachers said that they felt their classroom instruction (including texts, teaching, and assignments) was an "OK" or "Excellent" match. Approximately half of the urban teachers (47.8%) and exactly half of the rural teachers felt their practices matched this strongly.

In contrast, rural and urban teachers were more likely than suburban teachers to report that their students keep math portfolios—one of the hallmarks of innovative practice because portfolios are thought to encourage diversity of mathematics work, including math projects, writing, and investigations. Eighty-seven percent of urban teachers and 83.3% of rural teachers so reported, whereas only two (33.3%) of the suburban teachers reported having their students keep math portfolios; $\chi^2(2, N = 35) = 7.92$, $p = .02$.

Student preparation for concepts assessed on the CLAS. Students and teachers were asked to gauge the extent to which their classes had prepared students for some of the math concepts included on the CLAS—focusing particularly on the specific topic areas covered on the common CLAS form that was used for the retrospective student interviews. Students were asked how well prepared they thought they were for fractions, area, perimeter, graphing data, distance/time problems, and ratios. Similarly, teachers were asked how much class time was spent on these same areas. Students and teachers alike seem to agree that students were at least somewhat prepared in each of these areas, except for distance/time problems in urban classrooms and perimeter problems in suburban schools. While the patterns are somewhat irregular for teacher reports, for the most part students in the suburban schools tended to feel that they were better prepared in these content areas (see Tables 3 and 4).

Table 3

Percentage of Classes That Spent More Than Six Class Sessions on Content Areas by School Type (Teacher Interview Results)

Content area	Urban	Rural	Suburban
Fractions	73.9	6	100.0
Area	47.8	83.3	50.0
Perimeter	40.9	66.7	33.4
Graphing data	78.2	66.7	100.0
Distance/time	34.7	83.4	50.0
Proportional reasoning	73.9	66.7	83.3

Table 4

Mean Comparisons of Student Ratings of Their Preparation in Various Content Areas by School Type (Student Survey Results)

Content area	Urban	Rural	Suburban	$F(2)$
Fractions*	2.68	2.38	2.86	26.92
Area*	2.34	2.32	2.70	16.50
Perimeter*	2.25	2.41	2.72	22.60
Graphing data*	2.34	2.54	2.59	9.15
Distance/time*	2.16	2.28	2.40	5.69
Ratios*	2.11	1.85	2.58	30.65

Note. 1 = Little or none, 3 = Very well.
*$p < .05$.

Teaching and instructional strategies that build complex thinking. Alternative assessment is intended to emphasize open-ended problems that require not only a solution but also an explanation of how the student arrived at such a solution; the assessment thus values both complex mathematical thinking and communication. Both students and teachers were asked how often they engage in instructional practices that are associated with the development of these skills. Regarding these, a majority of students reported that they often solve word problems, solve problems that require thinking and that can be solved in more than one way, and use calculators (see Table 5). Students were less likely to

Table 5

Mean Comparisons of Student Frequency Ratings of Their Engagement in Specific Activities by School Type (Student Survey Results)

Activity	Urban	Rural	Suburban	Totals
Practice computations	4.96	4.33	5.60	4.96
Practice word problems	4.21	4.42	4.59	4.31
Problems solved more than one way	4.31	4.64	4.77	4.44
Problems that require you to really think	4.20	4.83	4.63	4.38
Problems where you explain your thinking	3.54	3.97	3.16	3.55
Problems that take at least a week to complete	2.21	3.17	1.58	2.27
Problems that apply to real life	3.48	3.68	3.58	3.53
Use calculators to solve problems	4.70	4.38	5.01	4.70
Use rulers, blocks, or solids	3.36	3.75	2.96	3.37
Give an oral presentation	2.42	2.99	1.82	2.42

Note. 1 = Hardly at all, 6 = A couple of times a week or more.

report working on problems for which they must explain their thinking; that take at least a week to complete; that reflect real-life problems; for which they use rulers, blocks or solids; or that require oral presentations. For comparison purposes, students also were asked how often they practice computations: Students in all classroom types, and particularly those in suburban classrooms, reported frequent engagement in such practice. Computation practice, in fact, in general was the highest frequency activity of all those queried. No significant differences were found across school type for any of these activities, except for solving problems that take at least a week to complete in which suburban students engage less; $F(2, 32) = 3.63$, $p = .038$.

Teachers' and students' responses regarding the frequency of another activity associated with innovative instructional practice, working in small groups, are shown in Tables 6 and 7. Students were consistently more conservative in their frequency estimates than were teachers, but according to the reports of both teachers, $\chi^2(10, N = 34) = 24.77$, $p = .006$, and students; $F(2, 32) = 7.80$, $p = .002$, it is clear that students in suburban classes were unlikely to be engaged regularly in small-group work, while rural students in our sample were most likely to be so engaged.

Table 6

Percentage of Teachers Who Reported Work in Small Groups at Least Once a Week by School Type (Teacher Interview Results)

Type of school	Teachers*
Urban	63.6
Rural	83.3
Suburban	0.0

*$p < .05$.

Table 7

Mean Comparisons of Student Frequency Ratings of Small-Group Work by School Type (Student Survey Results)

Type of school	Students*
Urban	3.40
Rural	4.77
Suburban	2.09

Note. 1 = Hardly at all, 6 = A couple of times a week or more.

We also asked students and teachers how often they worked on assignments that required extended writing (in the query to students, problems that required them to write a paragraph or more). Although student survey differences were not significant, teachers in urban and rural schools were more likely than teachers in suburban schools to report such activity, $\chi^2(10, N = 35) = 19.39, p = .04$ (Table 8 and Table 9).

Table 8

Percentage of Teachers Who Reported Working on Problems Requiring Writing at Least Once a Week by School (Teacher Interview Results)

Type of school	Teachers*
Urban	73.9
Rural	83.3
Suburban	66.7

*$p < .05$.

Table 9

Mean Comparisons of Student Frequency Ratings of Working of Problems Requiring Writing at Least Once a Week by School Type (Student Survey Results)

Type of school	Students*
Urban	3.40
Rural	4.77
Suburban	2.09

Note. 1 = Hardly at all, 6 = A couple of times a week or more.

Composite opportunity-to-learn scales. Based on a combination of factor analysis and theoretical assumptions, students' responses regarding specific classroom practices were combined into three overall scales. These provide a more reliable test of differences in students' opportunity to learn. The scales and the individual items that constitute them are reported in Table 10; also in Table 10 are the reliabilities (measures of how the scales hold together) associated with each scale, based on Cronbach's alpha (Cronbach, 1951).

Table 10

Classroom Learning Opportunity Scales: Composite Items and Reliabilities (Student Survey Results)

Mathematical communication scale (alpha = .66)	
	Problems which require you to explain your thinking
	Work in small groups
	Give an oral presentation
	Problems that require a written paragraph
Applied problem solving scale (alpha = .69)	
	Practice word problems
	Problems that can be solved in more than one way
	Problems that require you to really think
	Problems that take at least a week to complete
	Problems that apply to real life
	Use rulers, blocks or solids
Topic preparation scale (alpha = .76)	
	Perimeter
	Graphing data
	Distance/time
	Fractions
	Ratios
	Area

The "communication" scale is made up of items that indicate how often students practice problems that require them to communicate how they are thinking. The "applied" scale refers to how often students practice practical problems using applied methods or real-life perspectives. As indicated in the Methods section above, these scales correlated highly with results of independent analysis of teachers' instructional materials. The "preparation" scale focuses on how well prepared students felt for the specific math concepts required on the CLAS common form. Students in suburban schools engaged in less mathematical communication than urban and rural students; $F(2, 32) = 6.05$, $p = .006$ (see Table 11). Students in suburban schools felt better prepared than urban and rural students for selected mathematics concepts required on the common form assessment; $F(2, 771) = 28.73$, $p < .001$. Insignificant trends also support the possibility that suburban students practice computations more regularly than other students in our sample. ("Computation" denotes frequency of practice in computation, a single item unrelated to other subscales.)

Table 11

Opportunity-to-Learn Composite Scales: ANOVA Findings by School Type (Student Survey Results)

Type of school	Communication	Applied	Preparation	Computation
Urban	12.37	21.76	13.89	4.96
Rural	15.22	24.48	13.81	4.33
Suburban	9.25	22.12	15.86	5.60
Totals	12.32	22.29	14.32	4.96

Homework. Students' responses about the frequency with which they were assigned homework and the difficulty level of that homework provide a possible window into why suburban students tend to report themselves better prepared than other students in our sample (see Table 12). Time on homework presumably

Table 12

Mean Comparisons of Student Ratings Regarding Homework by School Type (Student Survey Results)

Type of school	How often homework is assigned[a]	How long it takes to finish homework[b]	How difficult homework is[c]
Urban	5.32	2.57	3.03
Rural	3.54	2.68	3.02
Suburban	6.58	2.37	3.18

[a] 1 = Never, 7 = Every night.
[b] 1 = 15 minutes, 5 = More than one hour.
[c] 1 = Very easy, 5 = Very difficult.

represents learning time and thus additional opportunity to learn. In this regard, suburban students reported being assigned math homework more often than did urban students, who in turn reported more homework than did rural students; $F(2, 32) = 6.61$, $p = .004$. Whereas suburban students reported having homework four to five nights a week on average and urban students reported having homework about three nights a week on average, rural students reported homework assignments only once or twice per week on average. No differences were found in the time students reported spending on each homework assignment (30 to 45 minutes on average) or in the difficulty level of that homework ("moderate" on average). The time in the context of frequency of homework, however, means that suburban students spend significantly more time per week engaged in mathematics than their urban or rural peers do.

Preparation for the CLAS

Students' perceptions of preparedness. Teachers and students also were queried about their direct preparation for the CLAS; how well students felt they were prepared for the CLAS is reported in Table 13. A one-way analysis of

Table 13

Mean Comparisons of Student Ratings of Their Preparation for the CLAS by School Type (Student Survey Results)

Type of school	Preparation for the CLAS
Urban	2.76
Rural	2.50
Suburban	3.23

Note. 1 = Not at all, 4 = Very much so.

variance indicated that suburban students were significantly more confident about their preparedness for the CLAS than urban students, who were more confident than rural students; $F(2, 780) = 29.77$, $p < .001$. It is possible that—having done well on previous standardized tests—students in suburban schools generally have more academic self-confidence than students in either rural or urban schools.

Teacher reports on direct preparation. Almost all teachers indicated that they engaged their students in specific activities to prepare students for the CLAS. The state provided schools with a "CLAS Mathematics Sampler" to acquaint teachers and students with the type and nature of assessment they would encounter on the CLAS. The Sampler, as the name implies, included sample problems, and teachers were free to assign and work through these problems with their students. The great majority of teachers interviewed (91%) had both seen the Sampler and used it to prepare their students for the assessment, although rural teachers appeared less likely than other teachers to have done so (see Table 14). On average, teachers reported devoting between three and five class periods

Table 14

Percentage of Teachers Using the CLAS Sampler by School Type (Teacher Interview Results)

Type of school	Used CLAS Sampler
Urban	95.7
Rural	66.7
Suburban	100.0

(median response) to practice with the Sampler, although it is worth noting that one third of our teacher respondents reported spending nine or more classroom periods in such efforts. In order to prepare their students for the assessment, a number of teachers mentioned giving students “a problem of the week” featuring problems they thought typified the new mathematics curriculum—problems that had no obvious solution, that could be solved in a number of ways, and/or that were drawn from real life.

How well teachers expect their students to do on the CLAS. Teachers were asked to estimate the percentage of their students that they expected to do well on the CLAS open-ended and multiple-choice items. For the most part, teachers tended to think that about half of their students would do well on the open-ended portion of the assessment, and that a slightly higher proportion would do well on the multiple-choice items. Suburban teachers held the highest expectations for their students’ performance on the multiple-choice items; $F(2, 32) = 4.85$, $p = .014$ (see Table 15). On average, suburban teachers, mirroring their students’ relative confidence, expected about 75% of their students to do well on the multiple-choice portion of the assessment. No significant differences were found by school type in teachers’ expectations of their students’ performance on open-ended items.

Table 15

Mean Comparisons of Teachers' Expectations of How Many of Their Students Will Perform Well on the CLAS by School Type (Teacher Interview Results)

Type of school	Open-ended	Multiple-choice
Urban	3.87	4.09
Rural	3.83	3.67
Suburban	4.17	5.00

Note. 1 = None, 6 = Almost all.

Observations related to preparedness: Use of time. Because the CLAS was not intended as a timed test, and in fact teachers were instructed to "make special arrangements for students who are still productively engaged at the end of 45 minutes, providing additional time for them to complete their work," observation of the time students actually spent on the assessment provided some indication of their engagement level and the ease with which they completed the assessment. In addition, because an open-ended item appeared first on the assessment, observers were able to note how long students spent on that item—or at least whether students spent the time assessment developers had estimated was required for a thoughtful response.

Distributions of students in observed classes using at least 15 minutes to complete the first open-ended item are shown in Table 16. In general, most

Table 16

Class Distributions: Percentage of Students Using at Least 15 Minutes to Answer First Open-Ended Problem by School Type (CLAS Administration Observation Results)

Type of school	A few students	About 25% of students	About 50% of students	About 75% of students	Almost all students
Urban	4.9	12.2	31.7	14.6	36.6
Rural	0	4.3	4.3	17.4	73.9
Suburban	0	0	0	0	100
Totals	2.4	7.1	16.7	11.9	61.9

students (75%-100%) spent at least 15 minutes on this item. However, results also indicate significant differences in schools serving different types of communities, $\chi^2(8, N = 84) = 28.92, p < .001$. Whereas in 100% of the suburban classrooms observed, almost all of the students used at least the allotted 15 minutes to answer the first open-ended problem, such extended concentration by most students was observed in only 37% of the urban classrooms. In nearly half the urban classrooms, 50% or more of the students were observed to have moved on earlier in the period. Students in rural classrooms more closely resembled the suburban students, with about three-quarters of the students using the full 15 minutes to answer the first problem.

Observers were asked also to estimate the percentage of students who completed the CLAS during the regular assessment period (see Table 17). Overall, in most classrooms, 75% to 100% of the students finished during the allotted time period. However, significant differences across school types were again found, $\chi^2(12, N = 78) = 31.87, p = .001$. In almost all suburban classrooms observed (94%), most or all students finished the assessment during the regular assessment period, while in only two-thirds of the urban classrooms did observers report that most students finished during this period. Rural schools showed the lowest completion rate, with most students finishing within the allotted time in only 35% of the rural classrooms observed.

Table 17

Class Distributions: Percentage of Students Completing the CLAS within the Regular Assessment Period by School Type (CLAS Administration Observation Results)

Type of school	None	A few students	About 25% of students	About 50% of students	About 75% of students	Almost all students
Urban	5.3	10.5	2.6	2.6	13.2	65.8
Rural	0	4.3	0	17.4	43.5	34.8
Suburban	0	0	0	0	0	94.1
Totals	2.6	6.4	1.3	6.4	19.2	62.8

Results regarding students' time usage during the assessment thus provide a portrait of difficulty in urban classrooms. That students in these classrooms did not spend the allotted time on the first open-ended problem suggests that they did not fully develop their responses to the open-ended task and perhaps were frustrated by it. At the same time, urban students were less likely than their suburban peers to complete the full assessment in the 45 minutes generally allocated to it. Where did urban students spend their time? Is it possible that these students spent much more time on the multiple-choice problems, indicating they had greater difficulty with these items than students in other schools? It is also possible that students revisited their responses to the open-ended tasks later on in the assessment period, and worked back and forth between the open-ended and multiple-choice items. Although it also is conceivable that teachers at the different school types reacted differently to the time constraint, influencing how comfortable students felt in continuing to work past the allotted time period, it does appear that students in rural and suburban schools had more efficient strategies for completing the assessment than did urban students.

Observations related to preparedness: Students' questions. The kinds of questions students have during the administration of a new assessment provide a window on difficulties they may be experiencing. While classroom observations indicated that students overall did not ask many questions during the CLAS (mean number of questions per classroom = 1.5), there were significant differences across school types; $F(2, 85) = 3.21$, $p = .045$, in both the frequency and types of questions asked (see Table 18). Suburban students asked the most questions and were far more likely than other groups to pose procedural questions, such as "Where do I work the problem out?" or "Where do I start the multiple-choice?"; $\chi^2(10, N = 86) = 25.01$, $p = .005$.

Contrary results were found regarding questions about a key term, "assumption," which appeared in the directions for the second open-ended task. Students in the suburban classrooms rarely raised questions about the meaning of this term, but in about 40% of urban and rural classrooms, students asked for clarification; $\chi^2(8, N = 86) = 15.43$, $p = .05$ (see Table 18). Clearly, since questions in general tended to be raised less often in rural and urban classrooms than in suburban ones, the different frequency on the "assumption" term cannot be attributed to students' propensities for asking questions. Rather, these findings seem to indicate that relative to suburban students, students in rural and urban schools are less familiar with an important concept in mathematical thinking and problem solving. Although this may be a problem of technical vocabulary, as opposed to underlying conceptual understanding, it is clear that some urban and rural students were at a disadvantage when solving the second open-ended task.

Table 18

Questions During the CLAS Administration by School Type (CLAS Administration Observation Results)

Type of school	Mean number of total questions	% Classes with procedural questions	% Classes with "assumption" questions
Urban	1.44	29.3	43.9
Rural	.83	39.1	39.1
Suburban	2.17	63.6	4.5

Interview data on approach to problem solving. Students' responses to think-aloud protocols provide additional light on the difficulties urban students faced. Data were coded for the type of reasoning students used when solving specific assessment tasks, and descriptive statistics were calculated. "Mathematics-based reasoning" was defined as that which utilized disciplinary concepts (rightly or wrongly) or strategic lines of reasoning based on mathematical thinking. For example, students who reasoned about a problem, "Well, the area of the larger square minus the smaller square should give you the shaded area and to get area from perimeter you. . . ." Within mathematics-based reasoning, we coded responses erroneous in the sense that students did not appropriately represent the problem or showed major misconceptions in their response to it. Accurate lines of reasoning were defined as responses that captured major aspects of the problem and followed reasonable mathematical thinking to solve them. In this category, however, students may have used inefficient and less elegant strategies and may have made errors in the application of concepts (e.g., formula for area or perimeter, computation errors, etc.). Random trial-and-error approaches, in contrast, were nonlogical from a mathematics perspective. For example, trying to play with the numbers in a problem to come up with an answer given in the multiple-choice alternatives:

> Well, first I tried to multiply them, but that wasn't an answer, then I thought about adding them but that didn't work either, so then I subtracted them which gave me 6 and then I divided by 2 because there were two of them and the number 3 was an answer.

These latter types of responses were combined with students who admitted guessing. For the open-ended problems, students overwhelmingly followed some mathematics-based reasoning approach (whether reasonable or erroneous) rather than using a trial-and-error or guessing approach: Only 4.3% of the students' responses were coded as guesses (see Table 19). Differences were also found

Table 19

Percentage of Students Using Guessing or Mathematics-Based Reasoning Approach on Open-Ended Items (Student Interview Results)

Type of school	Non-mathematical trial-and-error or guessing	Erroneous reasoning	Mathematical reasoning
Urban	2.6	56.4	41.0
Rural	7.7	50.0	42.3
Suburban	5.7	20.0	74.3
Totals	4.3	46.0	49.6

across school types, with suburban students more likely to use a correct line of mathematical reasoning; $\chi^2(4, N = 139) = 14.19$, $p = .007$. Urban students were far more likely than other groups to guess on the multiple-choice portion of the test, and far more likely to use a nonmathematical approach to selecting an answer; $\chi^2(4, N = 187) = 44.18$, $p < .001$ (see Table 20).

Table 20

Percentage of Students Using Guessing or Mathematics-Based Reasoning Approach on Multiple-choice Items (Student Interview Results)

Type of school	Non-mathematical trial-and-error or guessing	Erroneous reasoning	Mathematical reasoning
Urban	43.9	51.2	4.9
Rural	22.6	61.3	16.1
Suburban	15.2	36.4	48.5
Totals	35.3	50.3	14.4

Summary and Conclusions

While the findings of our study are far from conclusive, they do raise a number of important issues in providing to all students the opportunity to learn the complex mathematical thinking and problem-solving skills that are at the heart of new kinds of mathematics assessments. Our work similarly illustrates some of the difficulties of assessing students' opportunity to learn and the challenges of achieving meaningful educational reform. In reviewing our findings, readers will do well to keep in mind the volunteer status of our sample: That our schools volunteered to take on more testing and effort in a special pilot study for the state probably bespeaks their support and enthusiasm for the CLAS and the types of instruction and learning it emphasized. As a result, we suspect that our findings on opportunity to learn are likely to be rosier than they would be in a more typical sample.

Thus, perhaps it is not surprising that a majority of teachers in the sample already engage their students at least weekly in many of the instructional activities that the CLAS is intended to encourage: word problems, problems that can be solved in more than one way, problems that require extended writing, use of calculators, problems that require students to think critically, and small-group work. However, students perceive that other types of activities associated with a thinking curriculum are less prevalent. Problems in which students explain their thinking, oral representations, projects that take a week or more to complete, use of manipulatives, and real-life problems are less visible in the curriculum. As an additional indicator of routine classroom practice, work with computations was the most frequently occurring activity of those queried, and teachers clearly expected their students to do less well on innovative open-ended items than on traditional multiple-choice ones.

Given the equity impetus to our inquiry, it is encouraging that we did not find consistent differences across school types in students' opportunity to learn. Those differences that did emerge, however, represent teaching and learning issues of significant consequence. On the one hand, contrary to the fears of some, urban students in our sample were not limited to a meager "drill and kill" curriculum, and in fact they and their rural peers appeared more likely than suburban students to be engaged in constructivist instructional practices associated with recent curriculum reforms (Resnick & Klopfer, 1989). Nonetheless, students in urban classrooms were also more likely to have questions about a key concept in mathematical thinking, "assumption," and were less likely to have access to recent texts, raising questions about the depth of their preparation. Our suburban students clearly felt better prepared for the assessment, and observations and think-aloud protocol data strongly accentuate suburban students' preparedness relative to their urban peers.

While we did not find differences by school types in teachers' backgrounds, the educational background of our teacher sample does raise basic questions about their expertise in the mathematics they are expected to teach. Half of the sample have neither an undergraduate major or minor in mathematics, and in our sample, only 25% of our math teachers had actually majored in math or math education during their college work. One must wonder, therefore, where such teachers could have acquired the knowledge and sophisticated understanding of mathematics required to fully support their students' developing understanding: Constructivist practices require teachers who can pose appropriate problems, probe, and respond productively to their students' mathematical thinking and reasoning. Such teaching requires grounding in the discipline—in the concepts, principles, and ways of thinking mathematically. While heartening to see that a majority of teachers had participated recently in professional development for mathematics and mathematics education, one also wonders whether 40 hours or so of training is sufficient to provide teachers with the mathematical background and capacity they need to implement the reform agenda for mathematics education. Lacking solid understanding or other access to expertise, how can such teachers be expected to teach to rigorous mathematical standards?

That many teachers lack extensive background in mathematics, furthermore, intensifies the impact of insufficiencies in other resources. For example, when teachers themselves do not have abundant content knowledge, they become more dependent on texts for the content and activities of their instruction (Flanders, 1994), even though many instructional reformers advise against overreliance on texts in instruction (Farnsworth, 1992). That students in urban classrooms were less likely than those in rural and suburban classrooms to have recent mathematics texts, therefore, takes on added importance as an indicator of the opportunity to learn. Differences in homework frequency is another example of an isolated finding of differences among schools, but one with significant repercussions: Substantial differences in weekly time engaged in homework translates into very significant differences in time engaged in learning, applying, and practicing mathematics over the course of a year—assuming of course that homework assignments are meaningful and not trivial.

That urban students are clearly struggling despite data suggesting they are involved in innovative instructional practices raises questions about both the quality of our opportunity-to-learn measures and the quality of those instructional practices. Although, on one hand, the relationships between materials analysis and student reports are encouraging, differences between teacher and student reports suggest that some of the answers one gets depend on whom one asks. Teacher and student reports were generally consistent in relative differences across school types, but the absolute level and frequency of activity reported by these two sources often differed substantially. Whose views best represent reality and what other kinds of data are necessary to assure valid inferences about the opportunity to learn are serious methodological questions requiring additional attention.

At best, our current measures provide data on the relative presence or absence of various aspects of opportunity to learn and are silent on the quality of implementation. For example, how well are teachers involving students in cooperative group activity? When urban teachers engage their students in extended explanations, do students receive appropriate feedback that enables them to understand possible misconceptions and deepen their mathematics knowledge and thinking? Our current measures give precedence to form over substance. Given even the little data we have about students' performances, coupled with that on teachers' content background, we wonder whether current instructional practice similarly is more form than substance. Ultimately both our measures and, more importantly, the opportunities they seek to assess, must reflect the latter—not only opportunities, but high-quality opportunities that truly support student learning.

References

California Department of Education. (1989). *A question of thinking: A first look at students' performance on open-ended questions in mathematics*. Sacramento, CA: Author.

California Department of Education. (1992). *Mathematics framework for California public schools*. Sacramento, CA: Author.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Farnsworth, W. E. (1992, May). In teaching science, let the textbook support the classwork, not vice-versa. *Scientist, 6*(11), 12.

Flanders, J. R. (1994, May). Textbooks, teachers and the SIMS test. *Journal for Research in Mathematics Education, 25,* 260-278.

Herman, J. L., & Dorr-Bremme, D. (1983). Uses of testing in the schools: A national profile. *New Directions for Testing and Measurement, 19*, 7-17.

Herman, J. L., & Golan, S. (1991). *Effects of standardized testing on teachers and learning—Another look* (CSE Tech. Rep. No. 334). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Herman, J. L., Klein, D. C. D., Heath, T. M., & Wakai, S. T. (1994). *A first look: Are claims for alternative assessment holding up* (CSE Tech. Rep. No. 391). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Kellaghan, T., & Madaus, G. F. (1991). National testing: Lessons for America from Europe. *Educational Leadership, 49*(3), 87-93.

Madaus, G. F. (1991). The effects of important tests on students: Implications for a national examination system. *Phi Delta Kappan, 73*, 226-231.

Resnick, L. B., & Klopfer, L. E. (1989). Toward the thinking curriculum: An overview. In L. B. Resnick & L. E. Klopfer (Eds.), *Toward the thinking curriculum: Current cognitive research* (1989 Yearbook of the Association for Supervision and Curriculum Development) (pp. 1-18). Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Shepard, L. (1991). Will national tests improve student learning? *Phi Delta Kappan, 73*, 232-238.

Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice, 10*, 7-11.