Issues in Assessing English Language Learners' Opportunity to Learn Mathematics

CSE Report 633

Joan L. Herman & Jamal Abedi
CRESST/University of California, Los Angeles

June 2004

Center for the Study of Evaluation (CSE)
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 4.2 Validity of Assessment and Accommodations for English Language Learners
Joan L. Herman and Jamal Abedi, Project Directors, CRESST.

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences (IES), U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, the U.S. Department of Education, or the McDonnell Foundation.

ISSUES IN ASSESSING ENGLISH LANGUAGE LEARNERS' OPPORTUNITY TO LEARN MATHEMATICS[1]

Joan L. Herman and Jamal Abedi

**National Center for Research on Evaluation,
Standards, & Student Testing (CRESST)**

UCLA Graduate School of Education and Information Studies

Abstract

The Annual Yearly Progress (AYP) requirements of No Child Left Behind (NCLB) underscore both the mandate and the challenge of assuring that English Language Learners (ELL) achieve the same high standards of performance that are expected of their native English speaking peers. The intent indeed is laudable: states, districts, schools, and teachers must be accountable for the learning of their ELLs as are the students themselves. ELLs can no longer be invisible in the educational system, their learning needs must be met, and they too must make steady progress the goal of all students being judged proficient based on statewide testing by the year 2014. Already, however, NCLB results suggest a different reality: ELL subgroups *are* being left behind and schools and districts serving significant proportions of ELLs are less likely to meet their AYP goals and more likely to be subject to corrective action. Fairness demands that ELLs have equitable opportunity to learn (OTL) that upon which they are assessed, especially if those assessments carry significant future consequences. Moreover, if NCLB goals are to be met and achievement gaps reduced, schools must move beyond the performance only orientation of AYP to understand why results are as they are and how to improve them. OTL data can help to provide guidance in these areas and to acknowledge the reality that ELLs' learning is unlikely to improve unless and until students have more effective opportunities to attain expected performance standards. We view this study as an interesting beginning. It was conceived as a pilot, the results of which add fuel to the concern for and underscore some of the complexities of adequately measuring OTL, and we look forward to the full study involving a larger and more representative sample of teachers and classrooms and a more robust outcome measure.

The Annual Yearly Progress (AYP) requirements of No Child Left Behind (NCLB) underscore both the mandate and the challenge of assuring that English Language Learners (ELLs) achieve the same high standards of performance that are

[1] This paper was presented Friday, April 16, at the 2004 AERA Convention in the Opportunity to Learn Symposium.

expected of their native English speaking peers. The intent indeed is laudable: states, districts, schools, and teachers must be accountable for the learning of their ELLs as are the students themselves. ELLs can no longer be invisible in the educational system, their learning needs must be met, and they too must make steady progress the goal of all students being judged proficient based on statewide testing by the year 2014.

Already, however, NCLB results suggest a different reality: ELL subgroups *are* being left behind and schools and districts serving significant proportions of ELLs are less likely to meet their AYP goals and more likely to be subject to corrective action (EdSource, 2004; others). Research shows that economically disadvantaged and culturally diverse subgroups of the population, including ELLs, have had less access than other students to a challenging curriculum that would prepare them for success on today's standards (Guiton & Oakes, 1995; Wang, 1998). ELLs also are grossly over-represented in those failing to pass high school proficiency exams (see, for example, EdSource, 2003) and in danger of facing the dire consequences that accompany that the absence of a high school diploma.

The reasons for ELLs lagging behind are many, including the extra requirement these students face relative to their peers in acquiring academic English language proficiency, the constraints their language skills may impose on their ability to benefit from or have effective access to content instruction in English, and the confounding of their language ability with their subject matter competency when they are assessed in English (Abedi, 2004). Just as with their fully English proficient peers, opportunity to learn (OTL) looms large as a possible barrier to the success of ELLs (Herman, Klein, & Abedi, 2000).

Fairness demands that ELLs have equitable opportunity to learn (OTL) that upon which they are assessed, especially if those assessments carry significant future consequences (Baker, 1999). Moreover, if NCLB goals are to be met and achievement gaps reduced, schools must move beyond the performance only orientation of AYP to understand why results are as they are and how to improve them. OTL data can help to provide guidance in these areas and to acknowledge the reality that ELLs' learning is unlikely to improve unless and until students have more effective opportunities to attain expected performance standards. Are students being given such opportunities? What do effective opportunities look like? Absent data on OTL, policy makers will be missing critical evidence on which to base their decision-making and schools will be missing critical feedback. Just as with student performance data, information on ELLs' OTL can focus attention, stimulate schools' thinking about the strengths and weaknesses of their curriculum and course offerings, and encourage insights about priorities for professional development, materials acquisition, and resource allocations.

While there are innumerable other potential and important uses of opportunity to learn data—for example, in studies of the instructional sensitivity of tests (Baker, Linn, & Herman 2003), to condition test results (Muthen et al, 1995), or in research on curriculum-effective pedagogical strategies—this list may be sufficient to motivate the purpose of the current paper: to explore selected issues in the measurement of OTL, to provide preliminary findings on the relationships between ELL status and OTL, and to raise questions for future study. Note that this paper is based on a pilot study led by Dr. Jamal Abedi, my co-author, with support from his CRESST team: Mary Courtney, Seth Leon, and Jenny Kao. The work was conducted preparatory to a full research study investigating these same issues.

Related Research

Research on students' OTL has received modest attention in the literature, dating back to John Carroll's coining of the term in the early 1960s, when it denoted whether students had sufficient time and received adequate instruction to learn (Carroll, 1963; Tate, 2001). In the recent decades since, escalating demands for accountability and higher standards of student performance have led to renewed interest in the concept, encouraging researchers to expand conceptions beyond consideration of time to include the nature and quality of instruction and its pre-requisites (Burstein, 1993; Burstein et al., 1995; McDonnell, 1995; Burstein & McDonnell, 1993; Smithson, Porter, & Blank, 1995; Stevens, 1996; Brewer, D.J., & Stacz, C., 1996; Porter, A. C. 1991). However, while there has been attention to the definition of OTL and ways to potentially measure it, there has been relatively little consideration of the quality of the measures so developed and scarce attention to OTL for ELLs. Our study thus draws on research examining OTL for the general population and on specific teaching and learning issues relevant to ELLs.

OTL instrumentation

Four OTL variables have been prevalent in research: content coverage, content exposure, content emphasis, and quality of instructional delivery (Stevens, Wiltz, & Bailey, 1998). *Content coverage*, which is the most commonly used indicator for OTL, refers to the actual coverage of core curriculum topics specific to a particular grade level or subject area. *Content exposure* refers to the amount of time teachers allocate to covering the content. *Content emphasis* refers to the emphasis given to certain topics that are part of the core curriculum. The *quality of instructional delivery* refers to how coherently teachers present lessons that enable students to understand what is being taught. Other researchers have also included *instructional strategies* and *instructional resources*, which refer to both materials and teacher preparation (Herman, Klein, & Abedi, 2000).

Prominent in the measurement of OTL and the alignment of curriculum, instruction, and assessment, Porter (2002) noted three types of tools that have been used to measure content and address alignment in the past quarter century: surveys of teachers on the content of their instruction; content analyses of instructional materials; and alignment indices describing the degree of overlap in content between, for example, standards and assessment. Colker, Toyama, Trevisan, & Haertel's (2003) recent review notes additional methods that have been commonly applied. In addition to teacher/student surveys and analysis of instructional materials, Colker et al. noted the strengths and weaknesses of using teacher logs (Burstein & McDonnell, 1993; Harskamp & Suhre, 1994); classroom observation/taping (Muskin, 1990; Stigler, Gonzales, Takako, Knoll, & Serrano, 1999); analysis and ratings of class behaviors, teacher assignments (Aschbacher, 1994; Clare, 2002); and archival data. Other studies have used teacher interviews and instructional artifacts (Herman & Klein, 1996, 1997; Muskin, 1990; Wang, 1998), analysis of science notebooks and the comparison of teacher logs with students' notes (Ruiz-Primo, Li, Ayala, & Shavelson, 1999), or having teachers rate items or mathematics topics (Gamoran, Porter, Smithson, & White, 1997), a strategy adapted from the Second International Mathematics Study and applied in depth in the Third International Mathematics and Science Study (Schmidt & McKnight, 1995). For example, Yoon, Burstein, and Gold (1991) investigated content coverage by using a teacher questionnaire adopted from the Second International Mathematics Study. Teachers were asked to report on whether 96 topics were taught or not during two consecutive years, and whether topics were new, reviewed, or extended. Results indicated that some topics were more sensitive to content coverage than others.

Based on the literature, surveys have been the most common means of probing OTL (Collie-Patterson, 2000; Firestone, Camilli, Yurecko, Monfils, & Mayrowetz, 2000; Gamoran et al., 1997; McDonnell, Burstein, Ormseth, Catterall, & Moody, 1990; Muthen et al., 1995; Snow-Renner, 1998; Winfield, 1993; Wiley & Yoon, 1995; Yoon &

Resnick, 1998). Yet the validity of survey data continues to be suspect (Herman et al., 2000; Mayer, 1999; McDonnell et al., 1990). Moreover, as Colker et al. (2003) have noted, there has been a movement away from simple measures of instructional strategies towards greater concern over how instruction shapes cognitive demand. Consequently, recent research on instructional content also probes the level of cognitive demand (Porter, 2002).

OTL and ELLs

As noted in the introduction, studies have consistently found that ethnic minority students, including ELLs generally achieve poorly in mathematics (Gross, 1993; Kim & Hocevar, 1998). While language complexity may confound ELL students' ability to show what they know (Abedi, Lord, Hofstetter, & Baker, 2000; Abedi & Lord, 2001; Abedi, Leon, & Mirocha, 2003), available evidence suggests that OTL may be a contributing factor as well. Ethnic minority students have less exposure to content and their instruction tends to cover less content relative to non-minority students (Masini, 2001). Moreover, research shows a dramatic under-representation in higher-level math courses, and over-representation in lower level mathematics courses among ethnic minority students, which affects their OTL (Gross, 1993; Jones et al., 1986; Oakes, 1990). Gross (1993) noted that teachers of low-ability classes tend to emphasize drill and practice, rather than higher-thought processes, which are emphasized by teachers of high-ability courses. Gamoran, Porter, Smithson, and White (1997) found lower mathematics achievement among high school students in general track classes as compared to those in college-preparatory classes, implying that the practice of ability grouping and tracking denies students opportunities to learn. Such an impact could be further compounded for students who are ELLs.

However, research also shows positive examples. In a four-year project to locate and analyze schools with exemplary science and mathematics programs for middle school Limited English Proficiency (LEP) students, Minicucci (1996) found that these schools gave LEP students access to stimulating science and mathematics curricula with instruction in either the students' primary language or in English using sheltered language techniques.

Moreover, research suggests specific techniques that help ELLs acquire academic English proficiency and develop content knowledge. Williams (2001), for example, noted the substantial effects of students' academic English proficiency on their opportunities to learn. Williams made specific suggestions on how teachers can help their ELL students: draw connections between similar cognates in English and Spanish for Spanish-speaking students; use scaffolding with visual imagery; emphasize written skills as much as oral skills; read aloud everyday; avoid idioms; speak clearly; promote diversity; and avoid making assumptions about student understanding.

Methodology

Based on the research, we explored two complimentary approaches for exploring ELLs' opportunity to learn Algebra I, representing opposite ends of the cost continuum. At the most cost efficient end, we used surveys of teachers and students to address content coverage and to gather information about students' language background. We also piloted a preliminary observation schedule to get more detailed perspective on the nature of teacher-student interactions. In the section below, we first describe the Algebra I course which provides the context for the study and then describe our sample, instrumentation, and analysis strategies.

The Two-Year Algebra Course

In response to mandates that all Grade 8 students enroll in Algebra, the district in which the research was conducted has implemented a two-year algebra course that spans Grades 8 and 9. The Grade 8 portion offers a fertile OTL research situation because the program goal is to create a more equitable learning opportunity for Grade 8 students—both ELL and non-ELL—not ready for the pace of a 1-year algebra course. The two-year algebra course is also the math course in which the majority of Grade 8 ELL students are enrolled. Yet, it is not a "language minority ghetto." Most of the eighth grade population is enrolled in two-year algebra, and about half of the enrollees are non-ELL students, albeit often with a Spanish home language background. It is important to note that the teachers in the district use a standard text and have been given a schedule to follow so that the state standards are met in a timely manner. This almost "lock-step" schedule may help explain some of the OTL results.

Sample

The pilot study sampled three urban schools, which represented a range of socioeconomic status levels (SES). Within these schools, nine Algebra 1 teachers volunteered to participate and 24 of their classes represented the sample. A total of

602 Grade 8 students were included in the survey portion of the study.[2] In addition, 50 students and their algebra teacher from a fourth school participated in pre-pilot testing and observation. Moreover, nine classes, comprised of 271 students, were selected to participate in the observation phase of the study.

Demographic information. Of the total student sample, 54% were female and 46% were male and similar proportions were classified as English proficient and ELL respectively. Information gathered from the school's student data revealed that four levels of English language development (ELD) are represented, with the majority of the ELL students classified as ELD5, the level prior to reclassification as English proficient. A large majority of the sample qualified for free lunch and about 80% were of Hispanic descent.

The 24 classes were nearly evenly distributed among those in which ELLs were the clear majority (comprising more than 75% of the class), mixed classes, and those in which non-ELLs were the clear majority (ELLs less than 25% of the class). Average class size was about 26 students.

Achievement information. Table 1 shows the distribution of English language proficiency and performance on standardized tests of reading and mathematics. The data shows that all students in the sample scored below the national norm group on standardized tests of reading and mathematics. As might be expected, ELLs scored considerably lower than English proficient students (EO, IFEP, and FREP), and lower levels of English Language proficiency (lower ELD levels) were associated with lower performance.

When asked on the Student Background Questionnaire which languages they spoke before they started going to school, 72% of the students chose Spanish as being at least one of their home languages. As for their self-reported language proficiency, 58% said that they could understand the teacher's directions in English "very well" and 35% said "well."

[2] Fifteen students had to be deleted from the original sample because of incomplete data.

Table 1

Participants' Prior Math and Reading Performance on SAT9 by ELL Designation.

ELL Designation	SAT9 Reading			SAT9 Math		
	N	Mean	SD	N	Mean	SD
EO	89	38.83	23.86	86	38.50	22.91
IFEP	40	46.25	24.61	40	51.88	25.80
RFEP / RDS	193	33.38	17.94	191	42.50	22.56
ELD5 / P	194	15.37	12.08	193	23.95	16.02
ELD4	12	13.92	12.86	12	25.67	22.61
ELD3	11	5.36	11.64	11	4.65	6.62
ELD2	25	2.60	15.92	25	1.82	14.85
Total	564	26.63	33.98	558	21.03	22.83

Note: ELD1 students (*n*=2) were very likely exempt from taking the SAT9.

Participating teachers. The participating teachers' experience ranged from 2 to 11-plus-years in middle school with a continuum of training, credential, and educational backgrounds. Three of the nine participating teachers had earned a greater-than-temporary teaching credential and three others had education beyond a bachelor's degree.

OTL Measures

As noted above, our study included both survey and observation measures of OTL. Each of these is described below.

OTL surveys. Based on the content standards for the two-year algebra course, we identified 28 content areas that are supposed to be covered by teachers in the first semester of Grade 8, the time period of this study. In the teacher questionnaire, we listed the 28 content areas and asked teachers to indicate whether they taught those content areas. We listed the same 28 content areas in the student background questionnaire and asked students to indicate which of the content areas their class had covered. Responses from both teachers and students provided cross-validation data to examine the validity of responses by teachers and students. We computed the percent of exact agreement between teachers' and students' responses by counting the number of exact matches (if both teacher and student indicated that the topic was covered) divided by the total number of content areas (28) multiplied by 100. The percent of exact agreement for all students was 64.9 %. For ELLs, the percent of exact agreement was 54.1 % and for non-ELLs, the percent of exact agreement was 73.4 %.

To analyze the effect of OTL on achievement, we focused on 11 of the 28 content/skill areas that are represented in the algebra test items that we used for measuring students' algebra knowledge in this study (see description of outcome measure below). We computed the percent of exact agreement between teachers' and students' responses in these 11 content/skill areas. The percent of exact agreement for all students in these 11 areas was 69.5%. For ELL students, the percent of exact agreement was 62.8 % and for non-ELL students, the percent of exact agreement was 75.5%. It should be indicated at this point that in this pilot study we were not able to determine whether there are differences in agreement between ELLs and non-ELLs) within classes—i.e., whether ELLs may be less likely to report OTL specific topics than non ELLs in the same classes. There were not enough classes with sufficient numbers of both ELL and non-ELL students to make such a determination.

Because we believe that OTL, in the context of these Algebra I classes, is largely controlled by teachers for their classes as a whole (and mediated through a single textbook), we thought it most appropriate to consider classroom level measures of OTL. Within each class, OTL was measured by computing the mean student response to each content/skill area. (For example, "Combining Like Terms" would receive a score of 0.50 if 50% of the students in that particular class marked that they had studied it in Grade 8. Student content/skill area OTL scores could therefore range from *0* to *1*.) A total class-level OTL measure was computed as the sum of the scores of these 11 areas. Thus, the total class-level OTL measure for a class would be 11 if all of the students agreed that all 11 content/skill areas had been taught. So, our class-level OTL measure is a class-level variable and ranges from 0 to 11. The decision to use student reports rather than teacher reports had several motivations. SInce it was based on the responses of many students, as opposed to a single teacher, the student measure was considered more reliable. Moreover, the correlation between student-reported content-aligned OTL and teacher reported content-aligned OTL at the classroom level was high (.753), and the student measures showed a stronger relationship to student performance than did the teacher measure, providing additional evidence of validity. We provide additional data below in the results section on the issue of OTL's composition as a class and/or student variable.

In any event, the resulting student measure of classroom OTL appeared unidimensional, based on the high correlations between content associated with various text chapters and principal components analysis showing a single underlying factor. Among additional evidence of the validity of the measure, a comparison of OTL ratings related to content on the test and OTL ratings of non-tested content was telling. (Recall that we asked students to rate their OTL for all the topics of Algebra 1 and then derived from those ratings a measure of OTL for the 11 topics that were addressed on the test). Table 2 shows the correlations between OTL ratings and performance for both tested and non-tested OTL ratings.

Table 2

Correlations of Class-Level OTL With Class-Level Math Achievement (*n*=24)

	Math Score	Math SAT9	Algebra Grade
Student OTL on test content	.720**	.693**	.037
Student OTL other content, not tested	.197	.305	-.112
Teacher OTL	.533**	.460*	.321

**$p < 0.01$, two-tailed.

The data show that test-aligned OTL questions are more strongly correlated with math scores than non-tested OTL. For example, correlation between test-aligned OTL with the 19-item math test was .720, and with the SAT9 math test score, the correlation was .693 compared to correlations of .197 and 305, respectively for the 17 OTL content areas not assessed by the math test and SAT9 math test respectively. The results also showed that algebra grade point is not a good criterion in examining the effects of OTL on students' learning. As noted above, the correlation between teacher-reported, test-aligned OTL and math scores was lower that the correlation between student-reported test-aligned OTL and math scores.

Classroom observation protocols. As noted above, nine of the 24 classrooms were observed, including nine classrooms, taught by a total of seven teachers, and comprised of 224 students. Two protocols were used, requiring two observers in each classroom. The first observation protocol tracked teacher activities related to the use of instructional techniques that were likely to benefit the learning of ELLs. At the end of the observation, Observer #1 gave Likert-scale ratings to the teachers' use of language scaffolding, comprehensible speech, building on background knowledge, on-going assessment of student comprehension, and skills practice. In each observation, a teacher could receive a score between 1 (low) and 4 (high) for each attribute and a total score between 5 and 20. The scores for each class's two observations were averaged.

The second observation tracked individual level student activity. Classroom seat charts were used to tally each time any student engaged in the following:

1. Raised hand
2. Called on
3. Asked question
4. Worked alone
5. Group work
6. Unrelated
7. Demonstrates solution visually

The frequency of behavior for each classroom was then averaged over students by their ELL status and ELL and non-ELL students were compared on each of these seven behaviors The observers were not aware which students were designated ELL and which were non-ELL students, and their language proficiency was not apparent. We later matched the observed students with their ELL status.

While these measures (total number of behaviors) may be conceived as continuous variables, the number of students/classes observed was not large enough to do any meaningful analyses. These data thus should be considered illustrative only, particularly because of the limitations of one observer being able to fully track the actions of multiple students.

Achievement Measures

We defined achievement in math as the total score on a 20-item algebra test specially compiled for this study. In the interests of available resources—both time and budget—we used available items from NAEP and TIMMS that matched course goals to construct our measure. The algebra test contains 20 items that cover 11 of the 28 target content areas. One of the items was deleted due to technical problems. As an estimate of the reliability of the test, the internal consistency coefficient (alpha) was computed. For the entire test (all 19 items), the alpha was .604 (n=602). To examine the impact of linguistic complexity on the reliability of the algebra test, we grouped items into linguistically complex (9 items) and less linguistically complex (10 items). For the complex items, the alpha was .227 and for the non-complex items, the alpha was .562, with a difference of .335. The results of internal consistency analyses indicate that:

The entire test suffers from low reliability, which may be the result of several factors. Because the test measures a number of different content areas, it may well not be unidimensional. The relatively small number of test items may be another factor. The level of the test items' linguistic complexity may also have had a profound impact on the reliability of the test, consistent with our earlier studies that suggest that language factors may be a source of measurement error and may reduce the reliability of the tests (Abedi, Leon, & Mirocha, 2003).

In addition to the algebra test score, we also used the math subscale of the "SAT9 Math" as well as the students' algebra class grade as outcome measures. While the correlation between the math score and SAT9 Math was .577, that between the math score and Math Grade was .294, raising serious questions about the validity of grades as a consistent measure of performance.

English Language Proficiency Measures

Four measures of each student's English language proficiency level were collected: (a) SAT9 Reading score, (b) SAT9 Language score, (c) a Language Assessment Scale (LAS) fluency subscale score, and (d) a word recognition test score. For the Reading Proficiency Battery, the reliability coefficient (internal consistency) for the 10-item LAS frequency subscale was .697 (n=602) and for the word recognition items, the internal consistency was .943 (n=602).

In addition to these three scores, a composite score of all three was computed. Rather than a simple composite of the three English measures, we obtained a latent composite to control for measurement error due to the lack of a perfect correlation between the three measures. The latent composite was obtained through a confirmatory factor analytical approach. The three English language measures were used as the measured variables to create a latent variable.

Student Background Questionnaire

Students were asked to complete a survey about their language background and prior instruction. Students were asked about their country of birth, time in the

U.S. and in U.S. classes, and were asked to self assess their comprehension of their teachers, tests, and math tests, etc.

Information from this survey was used to create an individual measure of preparation, which was a latent composite derived from three groups of student background variables. Formed based on how they relate to prior opportunities to learn, these variables were: (1) prior OTL math content, (2) years in schools in the United States, and (3) access to learning resources.

Please see Abedi et al., 2004 for additional detail on all aspects of the study methodology and evidence on the reliability and validity of the measures.

Results

In the results section below, we examine the relationship between our OTL measures and student performance, for both ELL and non-ELL students. In exploring why results are as they are, we analyze the relationship between students' language proficiency, OTL and performance and that between preliminary data on teachers' ELL-relevant pedagogy and reported OTL. The latter findings raise questions about how OTL should be defined and to what extent issues of effective access should be considered.

Is There a Relationship Between our Measures of Classroom OTL and Student Performance?

Multiple regression analyses. Three multiple regression models were run to determine whether there was a relationship between classroom-level OTL and performance on our math measure: one for all students, one for ELLs, and the third for non-ELLs (See Tables 3-5). In each model, we controlled for prior math ability and prior student preparation to isolate current class OTL. The prior algebra grade and Math SAT9 score were used as proxies for prior math ability.

The data in Tables 3 through 5 suggest that, in all three models, OTL has significant contribution in predicting students' math performance. For Model 1 (all students, Table 3), among the three predictors, OTL was the second most important predictor. Similarly, in Model 2 (ELL students, Table 4) OTL was the second most powerful predictor of math performance. However, in Model 3 (for non-ELL students, Table 5) OTL had the lowest predictive power among the three predictors. Although the R-square is modest, the results suggest that OTL is a more determining factor in algebra achievement for ELL students than for the non-ELL group.

Table 3

Multiple Regression for all Students, Controlling for Prior Math Ability

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	3.47	0.58		6.02	.000
Math SAT9	0.05	0.01	0.42	9.84	.000
Prior Algebra Grade	0.37	0.09	0.16	4.28	.000
Class OTL (Content aligned to test)	0.33	0.07	0.19	4.46	.000
Student preparation	0.10	0.05	0.08	1.90	.058

Note. R-Square= .392, df=524; Math score is the dependent variable

Table 4

Multiple Regressions for ELL students, Controlling for Prior Math Ability

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	3.48	0.80		4.35	.000
Math SAT9	0.04	0.01	0.29	4.40	.000
Prior Algebra Grade	0.34	0.12	0.18	2.78	.006
Class OTL (Content aligned to test)	0.32	0.11	0.20	3.02	.003
Student preparation	0.07	0.07	0.07	1.00	.321

Note. R-Square= .227, df=227; Math score is the dependent variable

Table 5

Multiple Regression for Non-ELL Students, Controlling for Prior Math Ability

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	4.40	0.90		4.90	.000
Math SAT9	0.05	0.01	0.42	7.41	.000
Prior Algebra Grade	0.47	0.13	0.20	3.74	.000
Class OTL (Content aligned to test)	0.24	0.11	0.11	2.21	.028
Student preparation	0.04	0.10	0.02	0.46	.643

Note. R-Square= .355, df=297; Math score is the dependent variable.

HLM analyses. Given the nested nature of the data, one might argue that hierarchical linear modeling would be a more appropriate analysis than regression, and thus we constructed a two-level model, with student data as Level 1 variables and classroom data as Level 2 variables.

Table 6

Model 1: HLM Analysis, Classroom-Level OTL Measure as the Level 2 Variable

Fixed Effect	Coefficient	Standard Error	Approx. T-ratio	d.f.	P-value
For Intercept 1, B0					
Intercept 2, G00	2.50	0.72	3.45	22	0.003
Classroom OTL, G01	0.45	0.09	4.90	22	0.000
For Prior Algebra Grade Slope, B1					
Intercept 2, G10	0.45	0.11	4.23	23	0.000
For Math SAT9 Slope, B2					
Intercept 2, G20	0.05	0.01	5.94	23	0.000
For Student Preparation Factor Slope, B3					
Intercept 2, G30	0.08	0.07	1.10	23	0.285

Table 6 summarizes the result of analyses for Model 1 in which the classroom-level OTL measure is the Level 2 variable and proxies for student ability and background—students' math grade point, SAT9 Math, and OTL preparation factor—are Level 1 variables. The results show that the classroom-level OTL measure has significant effects on the outcome variable (student performance on the 19-item math test). Additionally, student-level math grades and SAT9 Math scores affected the outcome variables. However, after accounting for the classroom-level OTL measure, the student-level preparation / OTL factor had no significant effect. The result that classroom OTL significantly impacts math outcome scores, even when controlling for prior math ability, is consistent with the earlier regression analyses performed to answer Research Question 1.

Do ELLs and non-ELLs Have the Same Levels of OTL?

Descriptive analyses. Table 7 shows the descriptive statistics for the class-level OTL measure by ELL status. As indicated earlier, the class-level OTL measure ranges from 0 to 11, with 0 representing students' impression of no opportunity to learn the content / skill areas that represented in the algebra test questions and 11 suggesting opportunity to learn all 11 areas in the test. Therefore, the higher the class-level OTL measure, the more the students indicated they had opportunity to learn the concepts / skills. As data in Table 7 show, the mean OTL for all students is 8.43 (SD=2.13) out of a perfect score of 11, suggesting the students reported a fair level of opportunity to learn the topics that were the content of the test.

However, the mean OTL varies by language. The data in Table 7 suggest that non-ELL students had a higher level of opportunity to learn than the ELL students. The mean OTL for non-ELL students was 9.31 (SD=1.23) as compared to the mean OTL of 7.29 (SD=2.49) for ELL students. The difference between the class-level OTL measure across the ELL categories was highly significant ($t=13.02$, $df=600$, $p<.001$).

Table 7

Descriptive Statistics for Class-Level OTL Measure by ELL status

	Mean	N	SD
OTL all cases	8.43	602	2.13
OTL ELL students	7.29	263	2.49
OTL non-ELL students	9.31	339	1.23

It is important to note that the standard deviation for the ELL group (SD=2.49) is twice the standard deviation for the non-ELL group (SD=1.23). These data suggest that the ELL students were less consistent among themselves than the non-ELL group about whether or not they had opportunity to learn.

While our student observation findings, as noted above, should be considered very preliminary, they too suggest that ELLs have less access or engagement in OTL than their non-ELL peers. For example, non-ELL students raised their hands on average 1.39 times per class and were called on by the teacher about once per class, compared to ELL students raising their hands an average of .91 times and being called on and average of .55 times.

HLM analyses. As noted above, the nested nature of the data enables us to use HLM models to investigate the relationships among and between classroom- and individual-level variables. Model 1 was reported above, showing the relationship between student ability and background, class level OTL, and student performance. In Model 2, we included classroom-level ELL, the proportion of ELL students in a class, as an additional Level 2 variable. We included this variable to test whether the proportion of ELL students in the classroom might affect instruction and thus be confounded with our classroom level OTL measure

Table 8 presents this HLM analysis. The only difference between Model 1 and Model 2 was that in Model 2 we included an additional Level 2 variable, i.e., classroom-level ELL. The main reason for including this variable was that earlier studies suggested the proportion of ELL students in a classroom could affect instruction and assessment, and therefore, could be confounded with the classroom-level OTL measure.

As the data in Table 8 suggest, both the classroom-level OTL measure (t=2.07, p=.051) and classroom-level ELL (t=2.95, p=.008) as classroom-level variables are significant predictors of student math test scores, suggesting that the classroom-level OTL measure and proportion of ELLs in a classroom are both associated with student performance in math. Similar to the data presented for Model 1, students' grades and SAT9 scores are predictors of students' performance in the 19-item math test. Once again, after entering OTL as a classroom-level variable, the student-level preparation factor was not a strong predictor of the outcome variable (t=0.90, p=0.376).

Table 8

Model 2: Classroom-Level OTL Measure and Classroom-Level ELL as the Level 2 Variable

Fixed Effect	Coefficient	Standard Error	Approx. T-ratio	d.f.	P-value
For Intercept 1, B0					
Intercept 2, G00	4.97	0.96	5.18	21	0.000
Classroom OTL, G01	0.21	0.10	2.07	21	0.051
LEP Proportion, G02	-1.43	0.49	-2.95	21	0.008
For Prior Algebra Grade Slope, B1					
Intercept 2, G10	0.48	0.10	4.55	23	0.000
For Math SAT9 Slope, B2					
Intercept 2, G20	0.05	0.01	5.72	23	0.000
For Preparation / OTL Background Factor Slope, B3					
Intercept 2, G30	0.06	0.07	0.90	23	0.376

What Factors May Account for Differences in OTL for ELLs and non-ELLs?

We hypothesized that language proficiency itself might be contributing to OTL, in that students who are not fully proficient in English might have difficulty understanding and fully benefiting from textual materials and teachers' instruction. Such a relationship could play out in OTL in at least two ways: because of language issues, teachers in classes with higher proportions of ELLs might proceed through the curriculum at a slower pace, resulting in less OTL relative to the full set of topics addressed by the test, or ELLs may not perceive OTL, because effectively they have not been able to fully understand or benefit from curriculum and instruction, even though they have been exposed.

Multiple regression analyses. To examine the level of impact of language proficiency on OTL, we predicted the class-level OTL measure from three of the four English measures, the Word Recognition, the SAT9 Reading, and the LAS measures. (Since the SAT9 language score is highly correlated with the SAT9 reading score, and using both resulted in too many missing cases, we included the SAT9 reading score only.) We ran the regression model for the total group of students, for the ELL students, and for non-ELL students separately. Table 9 presents a summary of multiple regression results for all students, Table 10 presents the results for ELL students, and Table 11 shows the results for non-ELL students.

Table 9

Multiple Regression for All Students (Class-Level OTL Measure as Outcome Variable)

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Constant	4.418	.364		12.132	.000
SAT9 Reading	.017	.004	.167	3.810	.000
LAS	.308	.048	.290	6.451	.000
Word Recognition	.024	.007	.155	3.511	.000

Note. R-Square= .253, df=563

As the data in Table 9 suggest, over 25% of the variance of OTL (for both ELL and non-ELL students) is explained by English language test scores. While there is a substantial overlap between the three measures of English language, each of the measures has some unique and significant contribution to the model. Inspecting the data under the "Standardized Coefficient" column suggests that the LAS score has the highest level of contribution, higher than SAT9 Reading and the Word Recognition score. This may be due to the fact that LAS has better discrimination power for the ELLs.

Table 10

Multiple Regression for ELL Students

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Constant	3.289	.589		5.582	.000
SAT9 Reading	.026	.013	.129	1.974	.050
LAS	.276	.085	.226	3.226	.001
Word Recognition	.040	.012	.220	3.319	.001

Note. R-Square= .209, df=241

Table 11

Multiple Regression for Non-ELL Students

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Constant	7.786	.401		19.398	.000
SAT9 Reading	.003	.004	.050	.832	.406
LAS	.192	.045	.259	4.288	.000
Word Recognition	.004	.006	-.039	-.645	.519

Note. R-Square= .072, df=321

The findings of regression analyses for ELL students are summarized in Table 10 and those for non-ELLs are shown in Table 11. Results for ELL students are similar to those reported for all students, with the R-Square for the ELL model (.209) being slightly lower than the R-Square for the entire group of students (.253). Similar to the findings for the entire group of students, the LAS fluency measure was the most powerful among the three predictors for ELL students.

However, results for non-ELL students look considerably different. The R-Square for non-ELL students indicates that the proficiency measures explain only about 7% of the variance of OTL.

Student understanding of instruction. Results from the background survey further underscore the relationship between students' English language ability and OTL. Table 12 shows descriptive data demonstrating the relationship between students' self-reports of their ability to understand their teacher's directions (in English) and reported levels of OTL.

One can hypothesize that the higher the level of understanding directions in English, the more proficient a student is in English, and the more proficient in English, the more the student benefits from OTL. The results of our analyses presented in Table 12 support this hypothesis. Students who indicated that they understand directions in English "very well," had a substantially higher class-level OTL measure mean (M=8.74, SD=1.74) compared to those who believe they do not understand directions in English "well at all" (M=6.50, SD=1.58).

Table 12

Class-level OTL Measure Means by Understanding Directions in English

I can understand my teachers directions	N	Mean	SD
Very well	349	8.74	1.74
Well	214	8.07	2.53
Not well	19	7.58	2.85
Not well at all	10	6.50	1.58
Missing	10		
Total	602	8.43	2.13

Teachers' use of techniques to promote understanding for ELLs. One might also hypothesize that the more teachers employ pedagogical strategies thought to benefit ELL students' learning, the greater will be ELL students' effective access to OTL. While the number of teachers observed was too small to provide generalizable findings, the relationship between our OTL measure and teachers' use of pedagogical strategies is intriguing.

Table 13 shows the relationship between class-level OTL and observations of ELL pedagogical strategies, including teachers' use of language scaffolding, comprehensible speech, building on background knowledge, on-going assessment of student comprehension, and skills practice. (Recall that for each observation, a teacher could receive a score between 1 (low) and 4 (high) for the use of each strategy and a total score between 5 and 20. Scores were then averages across the two observations conducted). Despite the small number of classes observed, Table 13 consistently shows increased use of pedagogical strategies is associated with greater OTL for all five strategies.

Table 13

Classroom-Observer-Rated OTL Means by Class-Level OTL Measure

Class-level OTL Measure	Scaffolding	Comprehensible Speech	Building on Background	On-going Assessment	Skills Practice	Total
5 (*n*=1)	1.50	2.50	2.50	2.50	2.00	11.00
8 (*n*=2)	0.75	2.75	2.75	2.50	2.25	11.00
9 (*n*=4)	3.25	3.00	3.25	2.88	2.25	14.63
11 (*n*=2)	4.00	4.00	4.00	3.50	4.00	19.50
Total (*n*=9)	2.67	3.11	3.22	2.89	2.61	14.50

Summary and Conclusions

We started the presentation with concerns for the performance of ELLs, particularly in light of current assessment mandates, and questions about how OTL might help both to explain and to support policy and practice aimed at improving that performance. The results of this pilot study add fuel to the concern for and underscore some of the complexities of adequately measuring OTL.

The results showed that a relatively simple composite of student survey ratings showed reasonable measurement qualities. There was good consistency between student and teacher ratings and between students in a classroom. The measure had desirable characteristics of unidimensionality, although admittedly it concerned only one of many potential dimensions of OTL. Further the relationship between student performance and survey results for tested content and for content that was not tested offered additional evidence of validity. The relationship between individual- and classroom-level OTL measures and student performance supported the use of the classroom-level measure over the individual level one, a finding that was reinforced by subsequent substantive analyses.

The strong relationship the study found between classroom-level OTL and student performance is both a substantive finding of the study and additional evidence of the validity of the measure. Regression results suggested a stronger relationship between OTL and student performance for ELL students than for non-ELL students, suggesting that OTL indeed is a critical factor for ELL students.

In light of the strong relationship between OTL and student performance for ELL students, our findings with regard to the relationship between language status and classroom-level OTL give pause. Descriptive results suggest clear differences in OTL for ELL and non-ELL students in the study, and HLM results suggest that the proportion of ELL students and OTL have important effects on student performance, even after controlling for students' prior ability and background. Observation findings, while very preliminary, also suggest inequities in OTL for ELL and non-ELL students. These data suggest that current debates about bias in testing for ELL students ought to give at least as much attention to bias in OTL. Our data suggest that differential OTL may indeed play a role in the depressed performance of ELLs.

The relationship between language proficiency and OTL suggests a number of hypotheses about OTL are lower for ELLs than for non-ELLs. Our results indicate—not very surprisingly—that students' ability to understand instruction confounds their access to OTL, yet when teachers use appropriate pedagogical techniques, such barriers may be reduced.

While we were pleased with our results and are fully cognizant of its pilot status and limitations, we also see a number of study areas that were problematic, and these we think generalize to other studies as well. One issue is the quality of the outcome measures used to gauge student learning. For this study, we were careful to construct a measure that only included items measuring content that was supposed to have been covered in curriculum. While expediency limited how well we could specify the content and select appropriate items, the technical quality of the resulting instrument was disappointing. Apparently, in seeking alignment through publicly available measures, we gave up reliability. we suspect that in most studies the compromise is in the opposite direction—i.e., the measures are reliable, but their alignment with curriculum or specific OTL content is suspect.

In any event, we believe the sensitivity of the outcome measures used in studies of OTL is a major issue. The research requires high quality measures that are closely aligned with expected OTL. This in turn requires careful specification of content and equally careful selection or development of items. Further, while our study used an overall measure of learning, ideally one would want to be able to derive subscales to look at the relations between OTL in specific content areas and learning of those areas. Our study attempted analyses linked to individual items, but issues of stability and individual item quality hindered our success.

A second problematic issue is the meaning of OTL itself. What is a reasonable definition of OTL and how far down into enacted instruction should it really go? Should we define OTL as exposure—exposure to the "right" content, at the "right" level of cognitive complexity and using the "right" process? Should the concept credit intent—teachers trying to do it right and/or thinking that they are? Or to what extent should the concept really incorporate quality of teaching and learning activity and students' engagement in effective instruction? Our findings with regard to the relationship between language proficiency and OTL, coupled with our preliminary observation results, suggest the limits of only looking at exposure. Exposure clearly does *not* assure effective access to curriculum and appropriate opportunities to learn, but without such opportunities, can sufficient learning occur? At the same time, how deeply can we probe for effective access? We suspect the answer depends on purpose, and that most of us would like to delve more deeply than is possible on a regular basis.

Yet feasibility of measurement is an important issue. Much as we would like to get the data, available resources simply are not sufficient to regularly collect information on the quality of students' OTL for policy purposes. Further, it seems that if we got the measures of student learning "right"—e.g., if it was clear what was being measured so that teachers were clear on what to teach, and we knew that language demands did not overwhelm students' ability to show their knowledge and that results were sensitive to instruction—we would not have to worry so much about separate measures of OTL. We would know that low performance actually meant low OTL. Current policy is predicated on such knowledge, but we know there are critical missing pieces on all fronts: test content is not generally well specified, the linguistic complexity of items confounds students' ability to show what they know, and validity studies addressing instructional sensitivity are virtually absent from current practice. Yet unless and until we know that our measures are sensitive to instruction, how can we classify schools or teachers as ineffective on the basis of assessment results? Ultimately, rich measures of OTL may play their most important role in studies examining the instructional sensitivity of high profile tests.

Finally, we view this study as an interesting beginning. It was conceived as a pilot and we look forward to the full study involving a larger and more representative sample of teachers and classrooms and a more robust outcome measure. In the full study, we hope to more fully develop our qualitative measures, explore the within as well as the between classroom differences in OTL for ELL and non ELL students, and get a better handle on the sources of inequity in OTL for ELL students.

REFERENCES

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219-234.

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of student language background on content-based performance: Analyses of extant data* (CSE Tech. Rep. No. 603). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*(3), 16-26.

Aschbacher, P. R. (1994). Helping teachers develop and use alternative assessments: Barriers and facilitators, *Educational Policy, 8*(2) 202-223.

Baker, E. L. (1999, April). *Testing and assessment: A progress report*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, PQ, Canada.

Baker, E. L., Linn, R. L., & Herman, J. L. (2003, Spring). From the directors: Using Knowledge to Manage Immediate Results and Long-Term Value. *The CRESST Line,* 1-3.

Brewer, D.J., & Stacz, C. (1996). *Enhancing opportunity to learn measures in NCES data.* Santa Monica, CA: RAND.

Burstein, L. (1993). *Validating national curriculum indicators: A conceptual overview of the RAND/CRESST NSF Project.* Paper presented at annual meeting of the American Educational Research Association, Atlanta.

Burstein, L., McDonnell, L.M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). Validating national curriculum indicators. DRU-1086-NSF. Santa Monica, CA: RAND.

Carroll, J. B. (1963). A model for school learning. *Teachers College Record, 64*, 723-733.

Clare, L. (November, 2002). *Using Teachers' Assignments as an Indicator of Classroom Practice* (CSE Technical Report 532). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Colker, A. M., Toyama, Y., Trevisan, M., & Haertel, G. (2003, April). *Literature review of instructional sensitivity and opportunity to learn studies*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Collie-Patterson, J. M. (2000, November). *The effects of four selected components of opportunity to learn on mathematics achievement of grade 12 students in New Providence, Bahamas*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Bowling Green, KY.

EdSource. (2004). *No child left behind in California*? Palo Alto, CA: EdSource, Inc.

EdSource. (2003). *California's school accountability system under the federal no child left behind act (NCLB).* Retrieved April 2004 from http://www.edsource.org/edu_acc_ayp.cfm.

Firestone, W. A., Camilli, G., Yurecko, M., Monfils, L., & Mayrowetz, D. (2000). State standards, socio-fiscal context and opportunity to learn in New Jersey. *Educational Policy Analysis Archives, 8*(35), 1068-2341.

Gamoran, A., Porter, A. C., Smithson, J. & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325-338.

Gross, S. (1993). Early mathematics performance and achievement: results of a study within a large suburban school system. *Journal of Negro Education, 62*(3), 269-287.

Guiton, G., & Oakes, J. (1995). Opportunity to learn and conceptions of educational equality. *Educational Evaluation and Policy Analysis, 17*(3), 323-336.

Harskamp, E., & Suhre, E. (1994). Assessing the opportunity to learn mathematics. *Evaluation Review, 18*(5), 627-642.

Herman, J. L., & Klein, D. C. D. (1997). *Assessing opportunity to learn: A California example* (CSE Tech. Rep. No. 453). Los Angeles, CA: University of California, National Center for Research on Education, Standards, and Student Testing.

Herman, J. L., & Klein, D. C. D. (1996). Evaluating equity in alternative assessment: An illustration of opportunity-to-learn issues. *Journal of Educational Research, 89*(4), 246-256.

Herman, J. L., Klein, D. C. D., & Abedi, J. (2000). Assessing students' opportunity to learn: teacher and student perspectives. *Educational Measurement: Issues and Practice, 19*(4), 16-24.

Jones, L.V., Davenport, E.C., Bryson, A., Bekhuis, T., & Zwick, E. (1986). Mathematics and science test scores as related to courses taken in high school and other factors. *Journal of Educational Measurement, 23*(3), 197-208.

Kim, S., & Hocevar, D. (1998). Racial differences in eighth-grade mathematics: achievement and opportunity to learn. *Clearing House, 71*(3), 175-178.

Masini, B. R. (2001, April). *Race differences in exposure to algebra and geometry among U.A. eighth-grade students.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*(3), 305-322.

McDonnell, L. M., Burstein, L., Ormseth, T., Catterall, J., & Moody, D. (June, 1990) *Discovering what schools really teach: Designing improved coursework indicators.* CRESST UCLA//RAND Technical report. Santa Monica, CA: The RAND Corporation.

Minicucci, C. (1996). *Learning science and English: How school reform advances scientific learning for limited English proficient middle school students* (Educational Practice Report No. 17). Santa Cruz: University of California, National Center for Research on Cultural Diversity and Second Language Learning.

Muskin, C. (1990, April). *Academic work in high school history classes: Opportunity to learn in six schools.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Muthen, B., Huang, L., Jo, B., Khoo, S., Goff, G. N., Novak, J., & Shih, J. (1995). *Opportunity-to-learn effects on achievement: analytical aspects.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science* (Report No. NSF-R-3928). Washington, D.C.: National Science Foundation.

Porter, A. C. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis, 13*(1), 13-29.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3-14.

Ruiz-Primo, M. A., Li, M., Ayala, C., & Shavelson, R. J. (1999, March). *Student science journals and the evidence they provide: Classroom learning and opportunity to learn.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.

Schmidt, W. H., & McKnight, C. C. (1995). Surveying educational opportunity in mathematics and science: An international perspective. *Educational Evaluation and Policy Analysis, 17*(3), 337-354.

Smithson, J.L., Porter, A.C., & Blank, R.K. (1995). *Describing the enacted curriculum: Development and dissemination of opportunity to learn indicators in science education*. Washington, DC: Council of the Chief State School Officers.

Snow-Renner, R. (1998, April). *Mathematics assessment practices in Colorado classrooms: Implications about variations in capacity and students' opportunities to learn.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Stevens, F. I. (1996, April). *The need to expand the opportunity to learn conceptual framework: Should students, parents, and school resources be included?* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Stevens, F. I., Wiltz, L. & Bailey, M. (1998). *Teachers' evaluations of the sustainability of opportunity to learn (OTL) assessment strategies. A national survey of classroom teachers in large urban school districts.* Philadelphia, PA: Temple University, Laboratory for Student Success / Center for Research in Human Development and Education.

Stigler, J. W., Gonzales, P., Takako, K., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study*. Washington, DC: US Department of Education.

Tate, W. (2001). Science education as a civil right: Urban schools and opportunity-to-learn considerations. *Journal of Research in Science Teaching, 38*(9), 1015-1028.

Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis. 20*(3), 137-156.

Wiley, D. E., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System (CLAS). *Educational Evaluation and Policy Analysis, 17*(3), 355-370.

Williams, J. A. (2001). Classroom conversations: Opportunities to learn for ESL students in mainstream classrooms. *Reading Teacher, 54*(8), 750-757.

Winfield, L. F. (1993). Investigating test content and curriculum content overlap to assess opportunity to learn. *Journal of Negro Education, 62*(3), 288-310.

Yoon, B., Burstein, L., & Gold, K. (1991). *Assessing the content validity of teachers' reports of content coverage and its relationship to student achievement* (CSE Tech. Rep. No. 328). Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.

Yoon, B., & Resnick, L. B. (1998). *Instructional validity, opportunity to learn and equity: New standards examinations for the California mathematics renaissance* (CSE Tech. Rep. No. 484). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing