Consequences and Validity of Performance Assessment for English Learners: Assessing Opportunity to Learn (OTL) in Grade 6 Language Arts

CSE Report 635

Christy Kim Boscardin, Zenaida Aguirre-Muñoz,
Marjorie Chinen, Seth Leon, and Hye Sook Shin
National Center for Research on Evaluation, Standards,
and Student Testing (CRESST)/UCLA

July 2004

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

Project 4.3 Consequences and Validity of Assessment
Christy Kim Boscardin and Zenaida Aguirre-Muñoz, Project Directors, CRESST/UCLA

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, U.S. Department of Education.

CONSEQUENCES AND VALIDITY OF PERFORMANCE ASSESSMENT FOR ENGLISH LEARNERS: ASSESSING OPPORTUNITY TO LEARN (OTL) IN GRADE 6 LANGUAGE ARTS

Christy Kim Boscardin, Zenaida Aguirre-Muñoz, Marjorie Chinen, Seth Leon, and Hye Sook Shin

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

Abstract

In response to the growing achievement gap between English Learners (ELs) and non-ELs, standards-based instruction and assessment have been promulgated at the state and federal level. Yet, the consequences of standards-based assessment reforms for ELs have rarely been systematically studied. The work reported here represents the initial study of a 4-year research project with the purpose of investigating how the implementation of standards-based performance assessments and related instructional strategies influences the achievement of ELs. In this study, we were specifically interested in identifying the opportunity-to-learn (OTL) variables that positively impact student performance. We also investigated potential differences in the impact of OTL on performance between ELs and non-ELs.

Our study suggested that there are several factors contributing to students' performance on the Language Arts Performance Assignment (LAPA). At the student level, the analysis suggested that the greatest contributors to individual students' LAPA scores were performance on the Stanford 9 Language test, ethnicity, gender, and language proficiency status.

At the teacher level, we found that content coverage was significantly associated with student performance. The study showed that higher levels of content coverage in both writing and literary analyses were associated with higher performance for all students, including ELs. We also found differential impact of one OTL variable, content coverage–writing, on ELs' performance. This finding indicates that the gap between ELs and non-ELs increases as teacher reports of content coverage–writing increase.

Introduction

In 1994, a total of 3,184,696 English Learner (EL) students were enrolled in U.S. schools. In California, the Department of Education estimated about 25% of the total student population were EL students in 2001.[1] In response to the growing achievement gap between ELs and non-ELs, standards-based instruction and

[1] For more information, see http://www.ed-data.k12.ca.us

assessment have been promulgated at the state and federal level. Further, assessment programs are increasingly expected to include ELs, and they sanction schools that are not able to show improvement in EL achievement. Yet, the consequences of standards-based assessment reforms for ELs have rarely been systematically studied. Also, one of the presumptive benefits of standards-based performance assessments is that they encourage schools and teachers to address standards related to complex thinking and problem solving. Despite the intentions of such reforms, a critical issue for both EL and non-EL students is how effectively teachers can move from a superficial understanding of new standards-based content (i.e., in assessments) to a deep and serious implementation that will improve student learning.

The work reported here represents the initial study of a 4-year research project. The purpose of the larger study is to investigate the impact of the implementation of standards-based performance assessments on EL achievement including the identification of specific factors that influence ELs' success. Impact on achievement was investigated in the current study by examining the extent to which background characteristics and opportunity to learn (OTL) the content and skills targeted by the assessments contributed to higher performance of ELs. We expect this work to inform the design of assessment systems that best support the learning of ELs, including the training and implementation processes needed to support productive teacher change.

A Focus on Academic Language

Previous studies have shown significant achievement gaps between ELs and non-ELs (Cocking & Chipman, 1988). As the stakes of assessments increase, so do concerns among researchers regarding the essential instruction and learning EL students need to do well on these measures. These concerns stem from the growing evidence that academic success is associated with the acquisition of cognitive academic language. (See August & Hakuta, 1997, for a review of this literature.) Though researchers have found it difficult to reach some consensus on the definition of this construct, they generally refer to linguistic proficiencies required for subject matter learning (Stevens, Butler, & Castellon-Wellington, 2000; Wong Fillmore & Snow, 2000), which is usually devoid of contextual cues necessary for the English Learner (Cummins, 1981, 1984). Such decontextualization requires broad knowledge of "words, phraseology, grammar, and pragmatic conventions for expression, understanding and interpretation" of academic content (Wong Fillmore

& Snow, 2000, p. 20), as well as those linguistic features unique to single subject areas. (See Stevens et al., 2000, for a cogent review of the dominant theories of academic language.)

Implicit in state and national standards (Bailey & Butler, 2002), academic language is also present in both standardized tests that are used to judge achievement (Bailey, 2000) and in subject matter texts. However, despite widespread recognition of the need to increase the academic language proficiency of ELs, teachers in general are still ill prepared to provide instructional support in this area. Attention to academic language proficiency requires going beyond discussions of content to an analysis of the language used in the texts for rhetorical and aesthetic effect, necessitating an understanding of what Wong Fillmore and Snow (2000) refer to as educational linguistics. This knowledge base includes an understanding of (a) basic linguistics—including language structure, language use in educational settings, and basic linguistic analysis; (b) cultural diversity; (c) sociolinguistics—including language policies and politics that affect schools, language contact, and related topics; (d) language development with a focus on academic language development; and (e) second language learning and teaching.

Also, previous research in academic language suggests that the types of language or discourse required in an academic setting may be very different from the types of language and experiences of many EL students (Heath, 1983). ELs are often not provided with the opportunity to rehearse and develop their emerging academic language skills. The process of learning academic language requires much more time than that needed to learn language for interacting on a social level with English speakers. Ability with social language is usually developed within the first 2 years of arrival in an English-speaking setting; however, ability in the language needed for learning academic content may require 5 to 8 years to develop, or longer, depending on the age and prior educational background of the student (Collier, 1995; Cummins, 1981). Given the length of time that it takes to acquire academic language, ELs who enter the school system at the secondary level are particularly vulnerable to school failure because the cognitive demands of the curriculum at this level are higher.

We are reporting here on the first phase of a 4-year research project; this initial study does not examine the impact of academic language instruction on student performance. However, in subsequent documents, we will report on a more in-depth study of the opportunities students receive to enhance their academic

language proficiency and the relationship between such opportunities and performance on the standards-based performance assessment described below.

Performance Assessment Reform Effort

Since the fall of 1999 the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), in collaboration with a local school district, has developed performance assignments that support standards-based instruction and improved student attainment of the state standards. These Language Arts Performance Assignments (LAPAs), in combination with other elements of the school district's standards-based instructional program, are intended to serve the district's accountability and school improvement goals. The LAPAs were specifically designed to support the district's efforts to improve teaching and learning. In addition, these performance assignments are expected to serve as more sensitive outcome measures to assess the effects of anticipated districtwide professional development efforts.

More specifically, the LAPAs represent a performance-based approach to assessing students' understanding of subject matter content. This approach is consistent with CRESST's previous work,[2] which builds directly on what has been learned about effective teaching and learning. These assessments are designed to challenge students to construct their own responses to open-ended prompts about literary works, with the responses expected to involve a substantive integration of text-based information and the construction of reasonable and thoughtful interpretations about this information. Unlike many on-demand assessments, the LAPAs are designed to be similar to extended instructional activities or projects used regularly in classrooms. This assessment approach allows test users to assess knowledge that cannot be easily measured with multiple-choice questions or short constructed-response items. In this way, the LAPAs complement California's standardized tests to provide the district with a multiple-measures approach to monitoring student performance in language arts. The LAPAs were also designed to target important content that is representative of the domain and reflects meaningful California content standards for English language arts.

[2] For more information regarding the CRESST models, see the following: Linn, Baker, and Dunbar (1991), Baker, Aschbacher, Niemi, and Sato (1992), and Baker, Freeman, and Clayton (1991).

Opportunity to Learn

Opportunity to learn (OTL) is one of several important factors impacting student achievement. Previously used as a means to make valid cross-national comparisons in mathematics achievement (McDonnell, 1995), the notion of OTL has taken a more central role in American educational policy, particularly in response to the inequitable distribution of educational resources and access to knowledge documented by many researchers (e.g., Darling-Hammond, 1990, 1994; Gross, 1993; Jackson, 1982; Kozol, 2000; Oakes, 1985). This work has reported differences in resources such as level of funding and the physical condition of facilities, as well as differences in resources tied to teacher qualifications and classroom practices. The latter are believed to lead more directly to differences in access to knowledge (e.g., Gross, 1993; Oakes, 1985).

The evidence demonstrating that educational inputs vary greatly across schools coupled with the additional evidence linking these discrepancies to student achievement makes it necessary to develop OTL measures that can detect potential differences in the educational experiences of different groups of students that will aid in the interpretation of test scores. Porter (1991) outlined three main reasons for collecting such information: description of educational opportunities provided by schools, evaluation of school reform, and explanation of student achievement. Winfield (1993) offered similar reasons and pointed out that when performance assessments are used as the basis of reform, OTL information is crucial, given the greater cognitive demands of performance assessments.

Whatever the purposes to be served by OTL indicators, an OTL indicator system must be designed to provide systematic and comprehensive information regarding the factors that contribute to student achievement, as well as how policies are functioning and whether the assumptions underlying the policies are correct (Herman, Klein, & Abedi, 2000). It must have the potential to identify new problems, as well as to address old questions (Oakes, 1985). That is, it must provide consistent information over time and detect changes in instruction, as well as provide information that is useful for all stakeholders (Porter, 1991). In his model for OTL, Porter (1991) outlined educational inputs, processes, and outputs. Inputs include fiscal and other resources, general teacher quality (e.g., preservice training), student background, and parent and community norms. Processes include both the organizational characteristics of schooling, such as the quality of state and district standards, and the instructional characteristics of schooling, such as curriculum and

teaching quality. Finally, outputs as defined by Porter include achievement, participation, and attitudes and aspirations.

In line with this body of work, CRESST's survey approach to investigating the impact of OTL on student outcomes has proven to be useful when interpreting student scores (Baker et al., 1995; Boscardin, Stoker, Kim, Kim, & Aguirre-Muñoz, 2002). The main focus of this work is in the area of instructional characteristics—the content of what is taught and the actual pedagogical strategies that teachers employ to teach such content. While the operationalization of OTL utilized in the study reported here was influenced by the work of Porter, we aim to extend the classroom processes factor to include variables that are specifically relevant for English Learners, particularly exposure to, and the learning of, academic language. The findings reported here, however, present the work in the first phase of this 4-year study, which lays the groundwork for subsequent research.

Impact of OTL on Student Achievement

Methods for measuring what is taught have been relatively weak, due in part to the high cost related to obtaining accurate and reliable data on school process, typically collected through classroom observations. The alternative—survey instruments—also poses some challenges, particularly if used for high-stakes accountability purposes, due to the limitations of self-report data.[3] Despite these concerns, progress has been made in the measurement of OTL with substantial empirical evidence documenting the importance of OTL variables in explaining students' test scores (e.g., Boscardin et al., 2002; Brophy & Good, 1986; Fisher et al., 1980; Leinhardt, 1983; McDonnell, Burstein, Ormseth, Catterall, & Moody, 1990; Stevenson & Stigler, 1992). While early research utilized more traditional analysis designs, such as regression analysis (e.g., Leinhardt, Zigmond, & Cooley, 1981), later studies have used more sophisticated techniques to examine the impact of OTL on student test scores.

A study by Muthén, Kao, and Burstein (1991) analyzed OTL impact using a latent variable approach. In that study, OTL was defined as whether the instructional content needed to answer the test items was taught during the year the assessment was given or during the prior year. Muthén et al. found that students were more likely to respond correctly to an item if they had had the opportunity to

[3] Self-reported data may be subject to errors in recall and to overreporting or underreporting based on social desirability and errors from proxy reporting.

learn the tested concepts and skills. Students who had this opportunity during the year the assessment was given were most likely to respond correctly to an item. Though retrospective accounts of what was taught have their limitations, the study is important because of the innovative psychometric approach used to investigate the complexities of OTL, an issue that plagued previous research.

Researchers since then have used multilevel analyses for examining classroom processes, which more accurately reflect the multidimensionality of the classroom. Innovations in statistical methodology have resulted in the ability to investigate more comprehensive definitions of OTL. Wang (1998), for example, examined four dimensions of OTL: content coverage, content exposure, content emphasis, and quality of instructional delivery. Content coverage measured whether or not the core curriculum was covered. Content exposure reflected the time allowed for and devoted to instruction and the depth of the teaching provided. Content emphasis indicated whether topics were selected for instruction geared at lower level skills (e.g., rote memorization) or for instruction that emphasized higher order skills (e.g., problem solving). Finally, quality of instructional delivery included variables that reveal how classroom practices affect students' academic instruction. Wang found that OTL variables were significant predictors of both written and hands-on test scores. Further, OTL effects varied by test format. Specifically, content exposure was the most significant predictor of students' written test scores, whereas quality of instructional delivery was the most significant predictor of the hands-on test scores. Wang suggested that these findings provide evidence for examining OTL as a multidimensional construct, and therefore that various dimensions of "OTL should be measured simultaneously to properly document the OTL-achievement relation" (p. 137).

Other findings point to the need to examine achievement scores in light of both the instructional strategies to which students are exposed (i.e., OTL) and background factors that may also be associated with performance. For example, Saxe, Gearhart, and Seltzer (1999) investigated the impact of classroom practices on students' mathematics achievement by conducting classroom observations and collecting pre-and post-instruction achievement data. In addition to finding that differences in achievement could be attributed to differences in instruction, Saxe et al. found that the relationship between classroom practice and achievement differed depending on students' prior knowledge. For students with some understanding of the content prior to instruction, this relationship was linear. On the other hand, for

students without prior knowledge, the relation was nonlinear; but in classrooms that demonstrated instruction aligned to reform efforts, student achievement increased.

Differences in performance as well as OTL have repeatedly been associated with other student background characteristics such as language background, ethnicity, and gender. For example, Abedi, Leon, and Mirocha (2000) found that students' language proficiency was associated with their performance on NAEP in mathematics. In addition, Guiton and Oakes (1995), utilizing Second International Mathematics Study (SIMS) data, found that classes predominantly composed of White and Asian students had higher levels on all of their indicators of teacher quality (teacher experience, education, and assignment—proportion of teaching assignment devoted to math classes) than did mixed or predominantly minority classes, but the difference was not statistically significant. Correlations between new content coverage in five topic areas suggested that significantly more fraction and ratio subtopics were "introduced" rather than covered in depth as the minority composition of the classes increased, which was associated with teachers' lower expectations for predominantly minority classes in fractions and ratios than for White and Asian classes in those topics. As a result of such lower expectations, some classes received more low-level math content. Winfield (1991) found similar results. Further, Guiton and Oakes found that regardless of students' initial achievement level, those students who were placed in lower level courses showed smaller gains over time than students of comparable achievement who were placed in higher level courses.

This research demonstrates the critical role of the teacher in the achievement of students, and also justifies continued research to identify more comprehensive OTL constructs. Such research can substantially increase understanding of the variables that influence achievement of underperforming groups of students, such as ELs. At a minimum, OTL data should promote discussions around what curriculum is actually enacted, why, and whether existing practices are accomplishing what is intended (Guiton & Oakes, 1995). These discussions are likely to be most effective at the school or district level when conducted as part of an ongoing collective dialogue (Darling-Hammond, 1994; Porter, 1991). If these conversations are to occur, we need mechanisms for collecting and reporting such information back to schools and districts. The current study is expected to yield gains in this regard.

The following research questions are addressed using students' performance on the Grade 6 LAPA:

1. What is the impact of OTL on students' performance on a standards-based performance assessment in language arts? Does the magnitude of the impact vary depending on students' EL status?

2. What are the effects of students' language status and other background characteristics on students' OTL and performance on the performance assessment?

Based on our findings we hope to provide guidance for developing OTL instruments that are more sensitive to the types of experiences to which ELs need exposure and that may be unique to this population of students. This research should also lead to the identification of effective instructional strategies, as well as additional content and skills, that will particularly benefit the education and performance of EL students.

Method

Overview

This section describes the sample used in the analyses, the procedures and instruments used for data collection, and the reliability of local scores. Grade 6 was the focus of this investigation because of the persistent underperformance of students overall starting at the middle school level. This grade level was chosen as well because the number of students categorized as EL decreases dramatically after Grade 6.

Sample

District. The district from which the sample was obtained is mid-sized and situated in the Los Angeles metropolitan area. Total enrollment is approximately 15,000 students of whom about 1,500 are enrolled in Grade 6 and 22% are identified as English Learners. During the year that this study was conducted, 56% of ELs were enrolled in English language mainstream classrooms, and 31% were in structured English immersion classrooms. Further, 56% of teachers were fully credentialed, and the remaining teachers were working toward their credential with 15% participating in either a university, district, or pre-intern program (data retrieved from www.cde.ca.gov). The district average number of years teaching was 11.9, and 16% of teachers were either in their first or second year of teaching. Table 1 presents the percentage of sixth-grade students placed in each of the performance categories on the California Standardized Testing and Reporting (STAR) system for both English Only (EO) students and ELs. While most students at this grade level did not perform

Table 1

Percentage of Grade 6 EO and EL Students Placed in Each Performance Categories on STAR by Proficiency Status

Performance category	Proficiency status	
	% EO[a]	% EL[b]
Advanced	3	0
Proficient	15	4
Basic	42	33
Below Basic	21	35
Far Below Basic	19	28

Note. EO = English Only, EL = English Learner.

[a] $n = 1056$. [b] $n = 325$.

at the proficient level, far fewer ELs performed at the proficient and advanced levels than did EO students (4% and 18% respectively). Further, a greater percentage of ELs performed at the Below Basic and Far Below Basic levels than did EO students (63% for ELs and 40% for EO students).

Table 2 reports the 2001-2002 results of the California English Language Development Test (CELDT) scores for ELs.

Teachers. We sampled 27 teachers from seven schools. Of these 27 teachers, 34.6% had 4 to 7 years of teaching experience; 30.8% had taught for 11 years or more; 40.7% had majored in humanities/history; 38% had a master's degree; 0% had a Ph.D.; and 44% were credentialed.

Table 2

Percentage of Grade 6 EL Students Placed in Each of the Proficiency Levels on the California English Language Development Test (CELDT) ($n = 326$)

Proficiency level	Percentage
Advanced	3
Early Advanced	14
Intermediate	49
Early Intermediate	17
Beginning	16

Note. EL = English Learner.

Students. A total of 1,038 Grade 6 students completed the LAPA in the 2001-2002 academic year. Table 3 presents the number and percentage of students by key background variables (i.e., gender, language proficiency, ethnicity, lunch program participation, and parental education). Due to constraints in the sample size, students whose first language was other than English but who were identified as Initially Fluent English Proficient (IFEP) and students whose identification had been changed to Redesignated Fluent English Proficient (RFEP) were combined with English Only (EO) students (i.e., students whose home language was English) for some analyses. The English proficiency of the students in the IFEP and RFEP categories is closer to the proficiency of EO students and thus the three groups were combined. This combined group is referred to as non-ELs.

Table 3

Proportion of Students by Key Background Variable (n = 1038)

Variable	n	%
Gender		
Male	499	48
Female	539	52
Proficiency		
EL	219	21
RFEP	84	8
IFEP	118	12
EO	609	59
Ethnicity		
African American	495	49
Hispanic	511	51
Lunch		
No free lunch	53	5
Free lunch	941	95
Parental education		
Not a high school graduate	195	20
High school graduate	565	57
Some college	149	15
College graduate	74	7
Graduate school	8	1

Note. EL = English Learner; RFEP = Redesignated Fluent English Proficient; IFEP = Initially Fluent English Proficient; EO = English Only.

Procedure and Instruments

Student assessment. All sixth-grade students who had been in the district for at least one year completed the district's Language Arts Performance Assignment (LAPA), a curriculum-embedded performance assessment designed to assess students' understanding and skills in language arts. This assessment was modeled after previous CRESST work and is currently undergoing validation studies. In this assessment, students were asked to select a literary work that contains a heroic character and to describe the qualities of the heroic character in writing, citing detailed information from the literary work. Among the characteristics students could write about were physical and personality traits, thoughts and motivations, and relationships with other characters. Thus, students were expected to analyze the story beyond the surface features of the plot and to support all assertions about the text with accurate citations.

Students were also expected to go through the stages of the writing process with support from the teacher in the form of mini-lessons. If groups of students were having difficulty with elements of the writing assignment, the teacher was allowed to provide a short 15- to 20-minute lesson to help them get through the assignment. Assistance, however, did not include direct feedback, such as editorial suggestions or the teacher's interpretations of the text.

Students were given 5 to 10 hours of class time over the course of 1 to 2 weeks to complete the assessment.

Scoring sessions and rubric. Teachers at each school were trained to rate the LAPA tasks in scoring sessions led by the Performance Assignment Leader (PAL) at each school site. To prepare for the training, the PALs participated in trainer training sessions led by CRESST staff. The following is a description of the training model that was provided to the PALs. It was not possible to observe all of the school-level training sessions; anecdotal accounts indicate that there was some variation in how PALs elected to train teachers at the school sites. Each session was to begin with an introduction that included a general overview of the project in order to provide context for the work to be done. The introduction also provided an explanation of the purpose and goals of the scoring session. Following the overview, the agenda for the scoring session was presented. At this point raters were given the opportunity to review and discuss the performance tasks to be scored. The purpose of the discussion was to familiarize raters with the tasks. Then raters were presented with

the scoring rubric, which was a 4-point, focused, holistic rubric designed to focus judgments on the content of the response as opposed to mechanics or grammar (a copy of the rubric can be found in Appendix A). That is, literary analysis—in this case characterization—is emphasized over punctuation or proper paragraph construction. Teachers were to base their judgments on (a) the extent to which students focused on important character features, (b) the level of support for assertions, (c) the overall coherence of the response, and (d) mechanical errors.

After review of the scoring rubric, a set of anchor papers were presented to the teachers. This set of three papers illustrates the qualities described by the rubric for each score point. The papers represent the lowest possible performance for each of score points 2, 3, and 4. An anchor paper for a score of 1 is not necessary because responses that demonstrate less knowledge and skills than the anchor score 2 paper are given a score of 1. After review of the anchor papers, teachers were presented with a set of eight additional papers on which to practice the application of the rubric. This process should occur in sets of two to three papers, initially starting with a set of two papers and then, as agreement increases, scoring sets of three papers. The goal of the training was to reach 70% agreement of the set of eight papers.

Reliability of LAPA scores. To determine the reliability of the LAPA scores, a sample of papers were randomly selected from each school and scored again by the district raters. The district raters received training from CRESST researchers. For the purpose of the reliability analysis, it was assumed that the scores provided by the district raters, called *central scores,* would be a more reliable indicator of students' performance on the LAPA than the scores provided by the individual teachers, called *local scores*. We intended to run the analyses on the central scores. However, this set of scores presented sample size problems when investigating teacher-level effects on student scores. Therefore, the scores obtained from teachers at the local schools were used in subsequent analyses.

A comparison between these two sets of scores provided information about the reliability of the local scores. Table 4 presents two indicators of reliability between the central (district) scores and the local scores by school, and the average across the seven schools. The first reliability indicator is the percentage of agreement at the cut score for proficiency (score 3), which we refer to as the reliability of the proficiency decision. The second reliability indicator is based on the percentage of exact score agreement within 1 score point (i.e., ± 1 score point). Agreement for the proficiency decision ranged from about 62% to about 82%. As expected, agreement within 1

Table 4

Percentage of Exact Score and Proficiency Agreement Levels Between District and Local Scores

School	% Agreement for proficiency	% Agreement within 1 score point
1	72.4	89.7
2	80.0	90.0
3	81.6	95.9
4	64.0	74.0
5	74.5	95.7
6	68.3	81.7
7	62.1	79.3
Average	71.8	86.6

score point was higher and ranged from 74% to about 96%. The average agreement level was about 72% for the proficiency decision and about 87% for agreement within 1 score point.

The relatively low agreement levels for the proficiency decision are cause for concern and can be attributable to factors such as problematic training materials, inadequate training, and teachers' unwillingness to use the criteria in the rubric. Anecdotal information about the school site training indicated that both poor training and teacher unresponsiveness may have played a role in the low agreement levels.

Despite this limitation of the data, the district intends to use the local school site scores for monitoring the impact of reform and the achievement of standards, and for reclassification of English proficiency status. Therefore, it is reasonable to use this data set to examine the impact of OTL on student performance on the LAPA and to investigate potential differential impact.

Teacher Opportunity to Learn Survey. Teachers who administered the LAPA assessment also completed a teacher survey intended to capture critical aspects of OTL. The survey (see Appendix B) contains six sections: teaching experience, teacher expertise in content topics, content coverage, classroom processes, assessment practices and assessment preparation, and classroom resources. A brief description of these six areas follows.

Teaching experience questions targeted information about the teaching experience of participating teachers. Specifically, teachers were asked to report on their total number of years teaching, total number of years teaching the course, and total number of years teaching at the school.

The *teacher expertise* component of the survey targeted information about teachers' education and preparation in the course content and pedagogy, as well as their expertise in assessment-specific content. With respect to education and preparation, teachers were asked to report the number of courses completed in both their undergraduate and graduate programs that were directly related to the content of the assessment. They were also asked to indicate whether they had completed a master's degree and to state specifically in what area this degree was granted. The thought here was that teachers with a master's degree in English literature would be better prepared to teach language arts than teachers without a master's degree or with one in an unrelated field. To obtain a more complete picture of the extent of the teachers' training, they were also asked to list recent professional development related to the language arts. The second set of questions related to teacher expertise involved knowledge of content topics specifically targeted by the assessment, such as literary analysis.

Content coverage questions addressed the degree to which key assessment content was covered throughout the course of the school year. The aim here was to investigate the extent to which students had adequate opportunity to learn what was measured by the assessment. Content coverage items corresponded with the items students were asked to complete about the amount of time dedicated to key assessment content areas. However, student reports were not used in this analysis due to low internal consistency.

Classroom processes questions targeted instructional processes that were consistent with those elicited by the assessment (e.g., analyzing literary works), as well as those believed to engender deep levels of content understanding.

Assessment practices included students' experience with similar test item formats, as well as direct preparation for the assessment (e.g., analyzing literary works) and other factors that were likely to influence their performance.

Additional classroom resources was the final area that teachers were asked to report on, such as classroom libraries.

Teacher Opportunity to Learn Survey validation. Prior to studying the effect of opportunity to learn (OTL) on the LAPA scores, the technical quality of the items on the teacher survey was examined for those constructs for which such an analysis was appropriate. In the teacher survey for Grade 6, there were six OTL constructs that were examined. The content coverage and assessment practices scales were subdivided into two constructs: literary analysis and writing for the former, and assessment practices and LAPA preparation for the latter. Internal consistency was examined by Cronbach's alpha coefficient to assess the reliability of each construct. Also, confirmatory factor analysis (CFA) was conducted for further validation of the instrument because previous CRESST studies had already established these OTL constructs.

The six OTL constructs measured by teacher surveys were as follows:

- Expertise
- Content coverage–literary analysis
- Content coverage–writing
- Classroom practice
- Assessment practice
- LAPA preparation

Internal consistency for the six constructs ranged from .59 to .95. The reliability for expertise items was $\alpha = .95$. For content coverage–literary analysis ($\alpha = .92$) and content coverage–writing ($\alpha = .84$), the reliabilities were also very high. We found slightly lower reliabilities for classroom process ($\alpha = .78$) and LAPA preparation ($\alpha = .76$). We found the lowest reliability for the assessment practices scale ($\alpha = .59$). Considering the small number of items in each construct, these results confirm that, overall, the teacher OTL survey was highly reliable. The results are presented in Table 5.

Factor loadings (see Table 6) resulting from the CFA for expertise showed that all the items were highly associated with this factor (item 7a, as the first item in this construct, was constrained to be 1 for comparison purposes). This was consistent across all other factors. This pattern of results suggests that the items on the survey measure the factors (constructs) well. A check of the model fit indices also confirmed the appropriateness of these factor models (CFI = 1.0 and TLI = 1.0).

Table 5

Internal Consistency

Construct	*N*	Item	Alpha
Expertise	25	7a, 7b, 7c	.95
Content coverage–literary analysis	26	9a, 9b, 9c, 9d, 9e	.92
Content coverage–writing	27	9f, 9g, 9h	.84
Classroom processes	24	10e, 10f, 10g	.78
Assessment practices	25	11a, 11b, 11c, 11d, 11e, 11f	.59
LAPA preparation	25	12a, 12b, 12c, 12d, 12e	.78

Note. LAPA = Language Arts Performance Assignment.

Table 6

Confirmatory Factor Analysis Results

Construct	*N*	Item no.	Estimates	*SE*	Est./*SE*
Expertise	25	7a	1.000	0.000	0.000
		7b	1.028	0.025	40.427
		7c	0.986	0.023	42.008
Content coverage–literary analysis	26	9a	1.000	0.000	0.000
		9b	1.011	0.112	9.033
		9c	1.003	0.109	9.163
		9d	0.980	0.111	8.832
		9e	1.025	0.094	10.897
Content coverage–writing	27	9f	1.000	0.000	0.000
		9g	0.824	0.112	7.335
		9h	0.691	0.101	6.860
Classroom processes	24	10e	1.000	0.000	0.000
		10f	1.272	0.169	7.542
		10g	1.375	0.188	7.314
Assessment practices	24	11a	1.000	0.000	0.000
		11b	0.558	0.139	4.009
		11c	0.549	0.113	4.847
		11d	0.997	0.099	10.041
		11e	0.612	0.112	5.442
		11f	0.320	0.121	2.638
LAPA preparation	25	12a	1.000	0.000	0.000
		12b	1.037	0.178	5.813
		12c	1.140	0.186	6.120
		12d	1.343	0.143	9.385
		12e	1.295	0.152	8.517

Note. *SE* = standard error, LAPA = Language Arts Performance Assignment.

Additionally, all the items in the model were statistically significant. All the factor loadings had estimates that were twice the size of their standard error (SE). Again, these results suggest that the items on the survey appear to measure the constructs well (see Appendix C for detailed results of the CFA).

Table 7 shows the correlation among the six constructs. As expected, the relationship between most of the constructs was positive and significant. The strongest relationship was found between the two content coverage constructs (.704). The next strongest relationship was between classroom processes and assessment practices (.625). Other significant correlations were in the low to moderate range (1.66 to .566). Assessment practice was significantly related to all other factors, and LAPA preparation was significantly correlated with all factors except expertise.

Analyses

In this report, two separate analyses were conducted to address the research questions. First, to identify the factors (i.e., student background characteristics) that influence student performance on LAPA, ordinal logistic regression analyses were conducted. Second, ordinal logistic hierarchical linear modeling (HLM) was conducted to identify the OTL variables that contribute to student performance. Due to the small sample size at level 2 (n = 20), OTL factors were included in the analysis separately. Before the core analyses are presented, general descriptive information is reported.

Table 7

Correlations Among Factors

	EXPERT	CNTNTLIT	CONTNTWR	CLSSPRCSS	ASSESS	LAPAPREP
EXPERT	1.00					
CNTNTLIT	0.300*	1.00				
CNTNTWR	0.106	0.704**	1.00			
CLSSPRCSS	0.269*	0.364**	0.513**	1.00		
ASSESS	0.364**	0.387**	0.566**	0.625**	1.00	
LAPAPREP	0.065	0.166*	0.521**	0.398**	0.462**	1.00

Note. EXPERT = expertise; CNTNTLIT = Content coverage–literary analysis; CONTNTWR = Content coverage–writing; CLSSPRCSS = Classroom processes; ASSESS = Assessment practices; LAPAPREP = LAPA preparation.

*significant at α = .05. ** significant at α = .01.

Ordinal Logistic Regression

To maximize the data exploration, we examined the student-level data using ordinal logistic regression analyses. Once we identified the student-level factors that influenced student performance, we were able to explore the OTL variables that impact student performance separately using the logistic HLM models.

An ordinal logistic regression model is used to describe relationships between an ordinal response outcome and a set of independent variables. Since LAPA scores are ordinal variables with four categories or four response levels rather than a continuous variable, ordinal logistic regression is the most appropriate model. When examining an ordinal dependent variable, there are a number of options: (a) Treat it as a linear outcome and use Least Squares Regression; (b) dichotomize the variable and fit a binary model; or (c) fit an ordinal logistic model. Given that our dependent variable is ordinal, ordinal logistic regression is the most appropriate choice.

In ordinal logistic regression, since there are multiple categories, we label the outcomes as 1, 2, 3, and 4. The outcomes are ordered from high performance to low performance. The model is as follows:

$$\log it_j = \log\left(\frac{\sum_{k=1}^{j} \pi_k}{1 - \sum_{k=1}^{j} \pi_k}\right) = \alpha + \beta x$$

The ratio within the parenthesis expresses the odds-ratio. The odds-ratio represents the cumulative probability to score "j" in comparison to the complement of that probability. To express the outcome variable as a linear function of students' background variables, the log of the odds is used.

Often, exponentiated ordinal logistic regression coefficients are interpreted as odds-ratios. So given a one-unit increase in the covariate, it increases the ratio of the odds. However, in this report, we provide the probability (likelihood) rather than the odds-ratio for easier interpretation of the results.

Ordinal Logistic Hierarchical Linear Models

Results of the survey were then analyzed in concert with student performance results using two-level ordinal logistic hierarchical linear models (ordinal logistic-HLM). The factors influencing student performance occurred in the context of classrooms, which gave rise to multilevel data. Usually, students within the same

classroom are affected by similar factors such as teacher characteristics and educational resources, as well as the environment of the classroom. HLM models provided a systematic way to investigate how teachers in general, and specifically in relation to the OTL variables, influenced student outcomes and whether these variables had any differential impacts on EL performance after adjusting for student-level variables. Given the 4-point scale of the LAPA, we examined the relationship between LAPA scores and classroom differences characterized by OTL variables using ordinal logistic HLM.

The final HLM model specified in our study is as follows:

p_m: Prob. (outcome category=m)
p^*_m: Prob. (outcome category ≤ m) = $p_1 + p_2 + \ldots + p_m$ (therefore, $p^*_4 = 1$)
(* category 1 : the highest, category 4 : the lowest)

Level 1 Model

$$\log it(p^*_{1ij}) = \log\left(\frac{p^*_{1ij}}{1-p^*_{1ij}}\right) = \beta_{0j} + \beta_{1j}x_{ij}$$

$$\log it(p^*_{2ij}) = \log\left(\frac{p^*_{2ij}}{1-p^*_{2ij}}\right) = \beta_{0j} + \beta_{1j}x_{ij} + \delta_{2j}$$

Level 2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(OTL_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(OTL_j) + u_{1j}$$
$$\delta_{2j} = \delta_2 \text{ and, } \delta_{3j} = \delta_3$$

γ_{00} represents the adjusted grand mean logit level for the highest category, holding constant the OTL level. Thresholds δs are typically held constant across level 2 units. Therefore, mean intercept for category ≤ 2 becomes $\gamma_{00} + \delta_2$ and, for category ≤ 3, $\gamma_{00} + \delta_3$. γ_{01} shows the increment in the mean level caused by one unit change in OTL. γ_{10} captures the average slope of X_1. γ_{11} shows the increment in the slope of X_1 caused by one unit change in OTL. γ_{01} and γ_{11} are the key parameters of interest in this study since this captures the effect of the OTL variables.

Results

Descriptive Results

The overall mean score on LAPA was 2.33 on a 1- to 4-point scale. Table 8 displays descriptive information on the LAPA by key background variables. The information presented in Table 8 is the mean values without holding other predictors constant. Table 9 provides the distribution of LAPA scores across the four levels of performance (1, indicating low performance, to 4, indicating high performance).

Table 8

Descriptive Information on the LAPA by Key Background Variables

	N	Mean	*SD*
Total students	1038	2.33	0.95
Gender			
Male	499	2.15	0.93
Female	539	2.50	0.95
Proficiency			
EL	219	2.10	0.83
IFEP	118	2.61	.90
RFEP	84	3.01	.88
EO	609	2.27	.96
Ethnicity			
African American	495	2.22	0.94
Hispanic	511	2.43	0.95
Lunch			
No free lunch	53	2.66	1.09
Free lunch	941	2.33	0.94
Parental education			
Not a high school graduate	195	2.44	0.95
High school graduate	565	2.30	0.95
Some college	149	2.26	0.95
College graduate	74	2.65	0.90
Graduate school	8	3.00	0.76

Note. LAPA = Language Arts Performance Assignment; EL = English Learner; IFEP = Initially Fluent English Proficient; RFEP = Redesignated Fluent English Proficient; EO = English Only.

Table 9

LAPA Score Distribution

Score	*N*	%
1	213	20.5
2	411	39.6
3	270	26.0
4	144	13.9
Total	1038	100.0

The chi-square statistics (see Table 10) revealed that students' gender, language proficiency, ethnicity, and socio-economic status (SES) as indicated by Lunch Program participation, are all significantly related to students' performance on the LAPA. In order to determine the source of these statistically significant differences, more detailed analyses were conducted and are reported later in the report.

As part of an initial descriptive analysis, we also examined the relationship between students' background characteristics and the LAPA scores through the Spearman's Rho correlations. Table 11 shows the correlations among the LAPA scores, student background variables, and the SAT-9 language score. The correlation between the LAPA and SAT-9 language score was the highest among all the indicators (0.55). This moderate correlation is within what would be expected given the differences between test formats (i.e., constructed response versus multiple choice). Also, EL status was negatively correlated with parental education (r = –0.31). This finding indicates that, on average, EL students tended to have less

Table 10

Chi-Square: Student Background Characteristics and LAPA Performance

	Chi-square	*df*	*p*-value
Gender	35.67	3	0.00
Language proficiency	22.13	3	0.00
Ethnicity	18.57	3	0.00
Lunch	12.10	3	0.01
Parental education	18.44	12	0.10

Table 11

Spearman' s Rho Correlation

	1	2	3	4	5	6	7
1. LAPA score	—	0.18**	–0.12**	0.12**	–0.07*	0.00	0.55**
2. Female		—	0.01	–0.003	–0.03	–0.01	0.19**
3. EL			—	0.53**	0.10**	–0.31**	–0.13**
4. Hispanic				—	0.14**	–0.42**	0.11**
5. Free lunch					—	–0.16**	–0.01
6. Parent education						—	0.05
7. SAT-9 language							—

* = .05. ** = .01.

educated parents than non-ELs. Similarly, Hispanic students tended to have parents with lower education levels (r = –0.42) compared with African American students.

Ordinal Logistic Regression Results

In order to examine the relationship of the background variables and the outcome categories of LSCORE (LAPA scores ranging from 1 [lowest] to 4 [highest]), an ordinal logistic regression was performed for students in Grade 6. Background variables entered into the model included a scaled language SAT-9 percentile rank, ethnicity, parent education, language proficiency status, Free Lunch status (SES), and gender. Ethnicity categories were limited to African American and Hispanic students as there were not enough students of other ethnicities for meaningful comparisons. The four categories of parent education were parents with less than a high school education, parents who graduated from high school, parents that had attended some college, and parents who were college graduates. Students were coded into three language proficiency categories: English Only, IFEP/RFEP, and EL. The variable names and descriptions of the variables are presented in Table 12.

Results of the logistic regression analyses are shown in Table 13. The Wald statistics and the significance column in Table 13 indicate that the most important variables related to LSCORE categories were scaled SAT-9 Language percentile rank, ethnicity, gender, and language proficiency status. Coefficient estimates are in the form of log odds and can be difficult to interpret. To simplify the interpretation we have converted the log odds to expected probabilities of LSCORE categories for

Table 12

Variables in Logistic Regression

Variable name	Variable	Value
LSCORE	LAPA score (outcome variable)	Outcome variable, categorical, 1 to 4
FEMALE	Gender: female	Dichotomous, 1 = Female, 0 = Male
LUNCH	Free lunch	Dichotomous, 1 = Free lunch, 0 = not
PEDUCA2	Parent education	4-point scale: 1 = Not graduate high school, 4 = Postsecondary education
ETHNI2	Ethnicity/Hispanic	Dichotomous, 1 = Hispanic, 0 = African American
SCLANG	Language score	Continuous, Min = 0, Max = 10
LANGPR2	Language proficiency	Categorical, 1 = English Only, 2 = IFEP/RFEP, 3 = EL (English Learner)

Table 13

Ordinal Logistic Parameter Estimates

	Estimate	*SE*	Wald	*df*	Sig.
[LSCORE = 1]	–1.543	0.208	54.901	1	0.000
[LSCORE = 2]	0.875	0.205	18.317	1	0.000
[LSCORE = 3]	2.753	0.222	153.460	1	0.000
Scaled Language PR	0.418	0.027	241.546	1	0.000
African American	–0.774	0.211	13.435	1	0.000
Hispanic–reference	0.000	.	.	0	.
College graduate	0.212	0.272	0.608	1	0.435
Some college	–0.442	0.224	3.908	1	0.048
HS graduate	–0.138	0.173	0.640	1	0.424
< HS graduate–reference	0.000	.	.	0	.
EL	–0.820	0.237	11.954	1	0.001
IFEP/RFEP	–0.150	0.234	0.414	1	0.520
English Only–reference	0.000	.	.	0	.
No Free Lunch	0.718	0.280	6.576	1	0.010
Free Lunch–reference	0.000	.	.	0	.
Female	0.433	0.124	12.161	1	0.000
Male–reference	0.000	.		0	.

Note. Nagelkerke r^2 = .364.

the most important predictors. The expected probabilities for each of these predictors while holding the remaining background variables constant are presented in Figures 1, 2, and 3.

Figure 1 shows the expected probability for each category of LSCORE for male and female students. For the purposes of this example, the background variables other than gender were evaluated as the following constants: mean language SAT-9 percentile rank score, African American, English Only, not receiving free lunch, and parents who were not high school graduates. Male students in the evaluated background categories had an expected probability of 0.25 (1 chance in 4) of falling into the lowest LSCORE category. Female students were less likely to obtain an LSCORE in the lowest category as seen by their expected probability of 0.18. Conversely female students were more likely than their male counterparts to achieve a high LSCORE (3 or 4). This suggests that female students on average were performing better on the LAPA than male students.

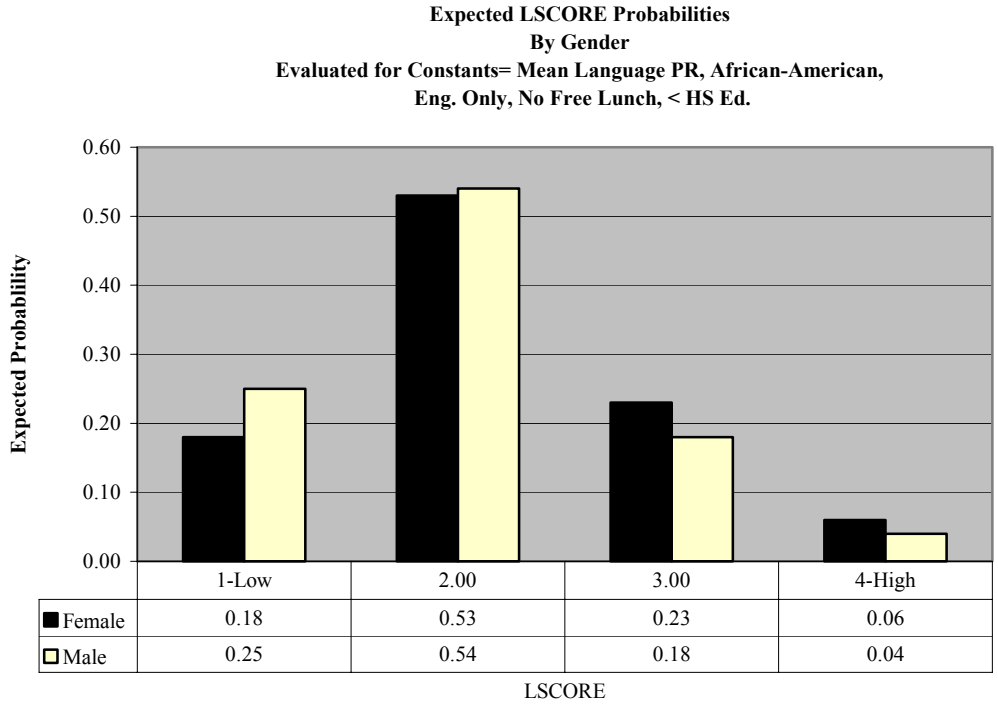


Figure 1. Expected LSCORE probabilities by gender.

Figure 2 shows the expected probability for each category of LSCORE by language proficiency status. Background variables other than language proficiency were evaluated as the following constants: mean language SAT-9 percentile rank, female, African American, not receiving Free Lunch, and parents who were not high school graduates. ELs in the evaluated background categories had an expected probability of 0.33 (1 chance in 3) of falling into the lowest LSCORE category. English Only and IFEP/RFEP students were less likely to obtain an LSCORE in the lowest category. The expected probability for these students was about 1 chance in 5. These results indicate that English Only and IFEP/RFEP students were more likely than ELs to achieve a high LSCORE (3 or 4).

Figure 3 shows the expected probability for each category of LSCORE by ethnicity. For the purposes of this example, the background variables other than language proficiency were evaluated as the following constants: mean language SAT-9 percentile rank, female, English Only, not receiving Free Lunch, and parents

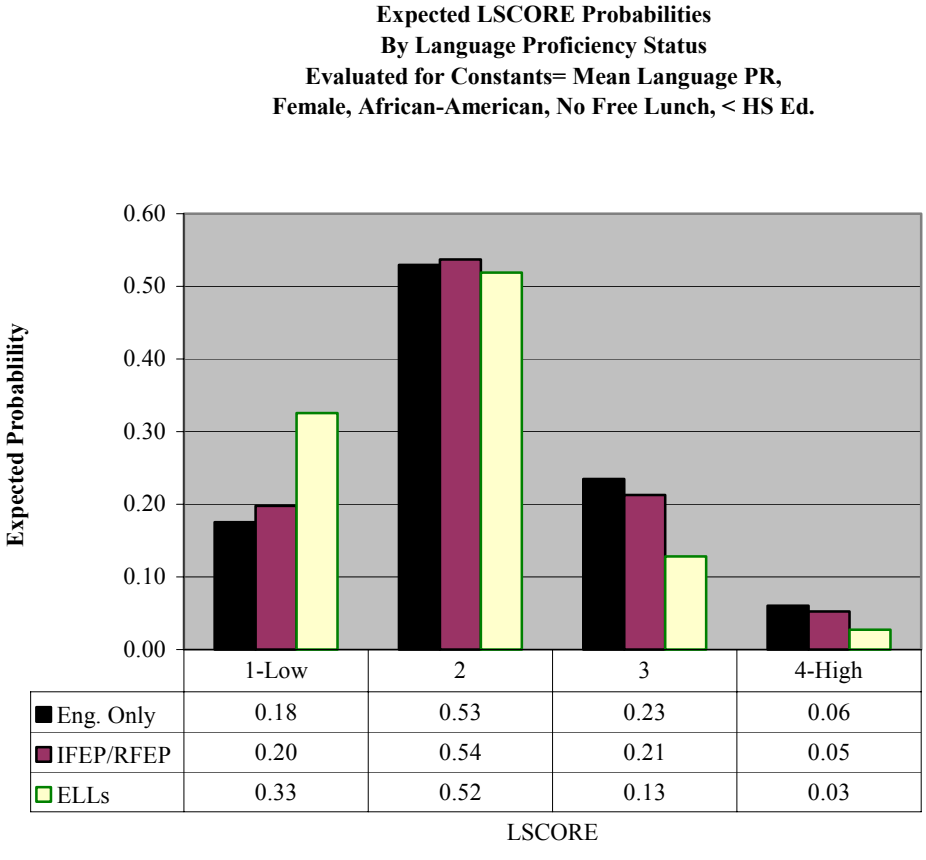


Figure 2. Expected LSCORE probabilities by language proficiency level.

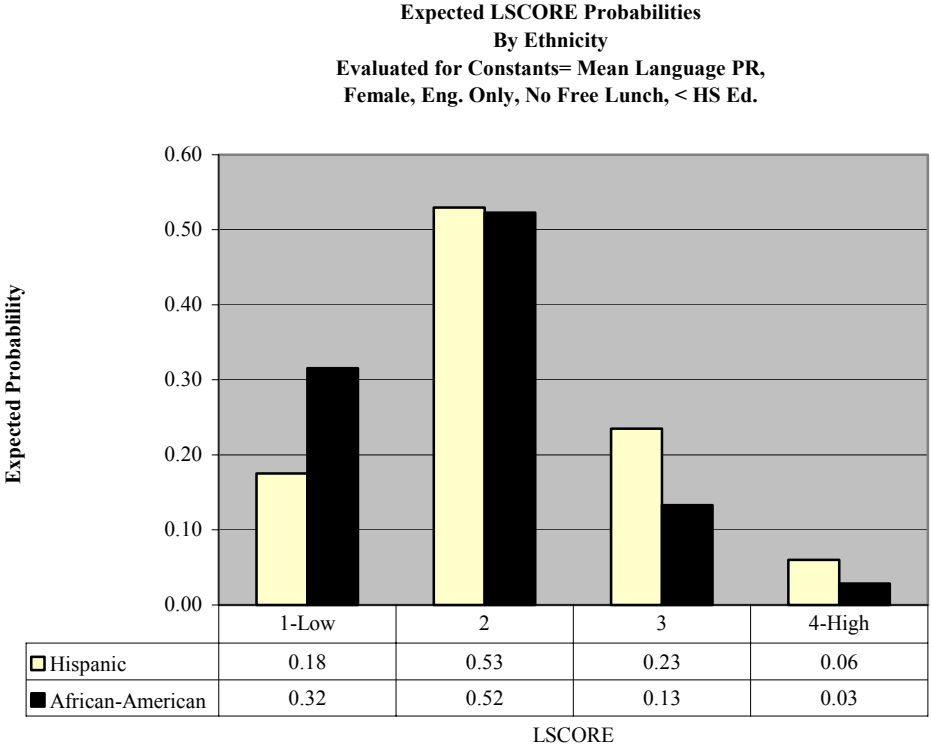


Figure 3. Expected LSCORE probabilities by ethnicity.

who were not high school graduates. African American students in the evaluated background categories had an expected probability of 0.32 (about 1 chance in 3) of falling into the lowest LSCORE category. Hispanic students were less likely to obtain an LSCORE in the lowest category. Their expected probability was less than 1 chance in 5. Similarly, Hispanic students were more likely than African American students to achieve a high LSCORE (3 or 4) while holding the remaining background variables constant. These results indicate that there was a statistically significant difference between African American and Hispanic students on LAPA performance.

Ordinal Logistic HLM Results

In order to determine the impact of OTL on student performance, ordinal logistic HLM analyses were performed. We also examined whether differential impact of OTL on language proficiency status was present. The outputs of these analyses are presented in Appendix D and are summarized below.

Level 1 Model

$$\text{logit}(p^*_{1ij}) = \log\left(\frac{p^*_{1ij}}{1-p^*_{1ij}}\right) = \beta_{0j} + \beta_{1j}(FEMALE_{ij}) + \beta_{2j}(LANG_{ij}) + \beta_{3j}(EL_{ij})$$

$$\text{logit}(p^*_{2ij}) = \log\left(\frac{p^*_{2ij}}{1-p^*_{2ij}}\right) = \beta_{0j} + \beta_{1j}(FEMALE_{ij}) + \beta_{2j}(LANG_{ij}) + \beta_{3j}(EL_{ij}) + \delta_{2j}$$

$$\text{logit}(p^*_{3ij}) = \log\left(\frac{p^*_{3ij}}{1-p^*_{3ij}}\right) = \beta_{0j} + \beta_{1j}(FEMALE_{ij}) + \beta_{2j}(LANG_{ij}) + \beta_{3j}(EL_{ij}) + \delta_{3j}$$

Level 2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(OTL_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31}(OTL_j)$$
$$\delta_{2j} = \delta_2 \text{ and, } \delta_{3j} = \delta_3$$

All the predictors (i.e., OTL constructs) are grand mean centered. Therefore, γ_{00} represents adjusted grand mean logit level for the highest category, holding constant OTL level. Thresholds δ s are typically held constant across level 2 units. Therefore, mean intercept for category ≤ 2 becomes $\gamma_{00} + \delta_2$ and, for category ≤ 3, $\gamma_{00} + \delta_3$. γ_{01} represents the incremental change in the mean level caused by one unit change in OTL. γ_{01} is the key parameter of interest since it captures the effect of OTL variable. γ_{10} captures the average gap between female and male students (female – male). γ_{20} is the expected change in outcome when LANG moves by one unit. γ_{30} shows the mean effect of ELs. γ_{31} indicates whether OTL leads to a larger or smaller EL effect. This provides information about the differential impact of OTL on ELs versus non-ELs. The variable names and descriptions are presented in Table 14.

This ordinal logistic HLM model was fit using nine teacher variables (including six OTL variables), one by one. Among the nine variables, we found that only two variables (CONTLIT and CONTWR) showed significant effects on student performance. The results are summarized in Table 15 (see also Table 17).

Table 14

Variables in Hierarchical Linear Models

Variable name	Variable	Value
Student-level		
LSCORE	LAPA score (outcome variable)	Categorical, 1 to 4
Gender	Gender: Female	Dichotomous, 1 = Female, 0 = Male
LANG	SAT-9 Language score	Continuous, Min = 0, Max = 10
EL	Language proficiency: EL	Dichotomous, 1 = EL, 0 = others
Teacher-level		
TYT	Total years of teaching	5-point scale, 1 = 1 year or less, 5 = 11 years or more
TECOURSE	No. English language arts courses taken	Continuous, Min = 2, Max = 12
CREDENT	Credential status	Dichotomous, 1 = Yes, 0 = No
EXPERT	Expertise	Continuous, Min = 1.00, Max = 3.00
CONTLIT	Content coverage–literary analysis	Continuous, Min = 1.00, Max = 5.00
CONTWR	Content coverage–writing	Continuous, Min = 1.00, Max = 5.00
CLSPRCSS	Classroom process	Continuous, Min = 1.33, Max = 4.67
ASSESS	Assessment	Continuous, Min = 2.50, Max = 4.33
LAPAPREP	LAPA preparation	Continuous, Min = 0.80, Max = 2.00

Table 15

Impact of CONTLIT (Content Coverage of Literary Analysis) on LAPA Performance

Fixed effects	Coefficient (*SE*)	*t*	*p*-value
For common intercept, β_{0j}			
Mean intercept, γ_{00}	–2.602 (0.226)	–11.511	0.000
CONTLIT effect, γ_{01}	**0.464 (0.194)**	**2.381**	**0.027**
Gender difference, γ_{10}	0.530 (0.140)	3.768	0.000
LANG effect, γ_{20}	0.480 (0.032)	14.719	0.000
For difference b/w EL and non EL, β_{3j}			
Mean EL difference, γ_{30}	–0.541 (0.194)	–2.781	0.006
Effect of CONTLIT on EL diff., γ_{31}	**–0.218 (0.161)**	**–1.350**	**0.177**
Threshold(2), δ_2	2.027 (0.124)	16.346	0.000
Threshold(3), δ_3	4.674 (0.183)	25.488	0.000

Given that the outcome is in a log-odds scale, γ_{00} captures the overall log-odds of getting the highest score (i.e., 4) after controlling for all the predictors at the grand mean level, including gender and EL status. However, this is very difficult to interpret. For easier interpretation, the log-odds were then transformed back to a probability scale as follows:

$$\text{Log}(p/1-p) = -2.602 \rightarrow p = \exp(-2.602)/(1 + \exp(-2.602) = 0.07.$$

Adding threshold(2) gives the probability of getting a score of 3 or 4 (i.e., passing the test).

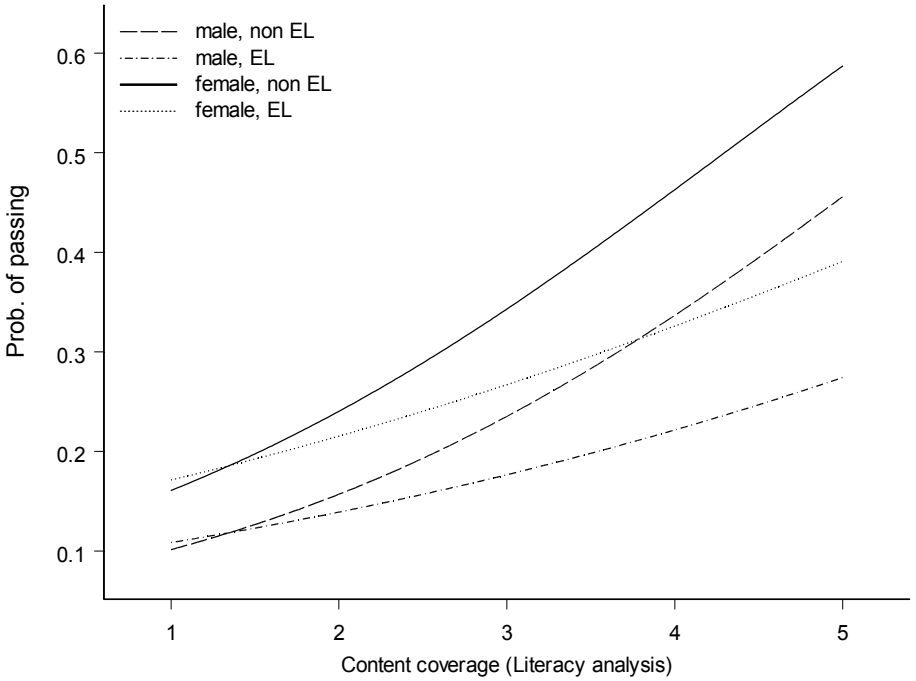
$$\text{Log}(p/1-p) = -2.602 + 2.027 = -.575 \rightarrow p = \exp(-.575)/(1 + \exp(-.575) = 0.36.$$

This means that students overall had about a 7% probability of getting the highest score and a probability of 36% of getting a score of 3 or higher after adjusting for GENDER, LANG, and EL.

CONTLIT, GENDER and LANG all had a positive effect on LSCORE. Therefore, these results suggest that females performed higher than males and that higher CONTLIT levels led to higher LAPA performance. In addition, students with higher scores on LANG also had a higher probability of obtaining higher LAPA performance.

EL status had a negative effect on LSCORE (–.541). In other words, ELs were more likely to perform lower on the LAPA than the non-ELs. The effect of CONTLIT on the EL effect was also negative (–.218). Although this effect was not significant, the direction of the pattern indicates that, holding constant GENDER and LANG, the gap between EL and non-EL students increases as the class CONTLIT level also increases (see Figure 4). The lack of significance could be attributable to the small number of teachers responding to the teacher questionnaire. Thus, whether or not content coverage–literary analysis differentially impacts EL performance is not clear.

Since all the variables were grand mean centered and the outcome was in the logit scale, for easier interpretation of the parameter estimates, students were categorized into several subgroups, and then the actual fitted probability for each group was calculated. Doing so allows for further exploration of potential of differential impact of OTL.



Note: The differential impact of CONTLIT on EL and non-EL is not significant.

Figure 4. Impact of CONTLIT on students' LSCORE by gender and language proficiency status.

Table 16 shows the various gender, EL, and CONTLIT groups' expected probability of getting a score of 3 or higher (passing this exam), holding constant LANG level at the grand mean = 1.49.

Table 16

Probability of Passing (Getting Score 3 or Higher, OTL = CONTLIT), LANG Fixed at 1.49

	CONTLIT level				
	1	2	3	4	5
Females					
EL	.17	.22	.27	.33	.39
Non EL	.16	.24	.34	.46	.59
Males					
EL	.11	.14	.18	.22	.27
Non EL	.10	.16	.24	.34	.46

Figure 4 illustrates several important findings. First, for all the students, as the content coverage increases, the probability of getting a high score on the LAPA also increases. Second, non-EL female students with high content coverage have the highest probability of getting a high score on the LAPA. Also, even though the gap between the performance of ELs vs. non-ELs is not statistically significant, this gap increases as the content coverage increases. Again, the nonsignificance finding is likely attributable to the small number of teachers who completed the survey.

Table 17 presents the results for the impact of content coverage–writing on LAPA performance. The outcome is also in a log-odds scale. Therefore, γ_{00} captures the log-odds of getting the highest score (i.e., score 4) controlling for all the predictors at the grand mean level, including gender and EL. The transformation back to probability is as follows:

$$\text{Log}(p/1–p) = –2.579 \rightarrow p = \exp(–2.579)/(1 + \exp(–2.579) = 0.07.$$

Adding threshold(2) gives the probability of getting score 3 or 4 (i.e., passing the test).

$$\text{Log}(p/1–p) = –2.579 + 2.033 = –.546 \rightarrow p = \exp(–.546)/(1 + \exp(–.546) = 0.37.$$

This means that, when adjusted for GENDER, LANG and EL, students have a 7% chance to get the highest score (4) and a 37% chance for a score of 3 or 4.

This indicates that females have a higher probability of receiving a hither score on the LAPA than males. Also, the results suggest that higher levels of CONTWR are associated with higher performance on the LAPA. In addition, students with higher scores on LANG also had a higher probability of obtaining higher LAPA performance.

We categorized students into several subgroups and calculated the actual fitted percentage for easier interpretation of the OTL effect on LSCORE (LAPA). Table 18 presents the expected probability of obtaining an LAPA score of 3 or 4 (passing this exam) for the various variables: GENDER, EL and CONTLIT groups, holding LANG level constant at the grand mean = 1.49.

Table 17

The Impact of CONTWR (Content Coverage of Writing) on LAPA Performance

Fixed effects	Coefficient (*SE*)	*t*	*p*-value
For common intercept, β_{0j}			
Mean intercept, γ_{00}	–2.579 (0.229)	–11.253	0.000
CONTWR effect, γ_{01}	**0.543 (0.222)**	**2.437**	**0.024**
Gender difference, γ_{10}	0.533 (0.140)	3.791	0.000
LANG effect, γ_{20}	0.486 (0.032)	14.858	0.000
For difference b/w EL and non-EL, β_{3j}			
Mean EL difference, γ_{30}	–0.543 (0.194)	–2.788	0.006
Effect of CONTWR on EL diff., γ_{31}	**–0.332 (0.193)**	**–1.713**	**0.086**
Threshold(2), δ_2	2.033 (0.124)	16.367	0.000
Threshold(3), δ_3	4.685 (0.183)	25.486	0.000

Table 18

Probability of Getting Score of 3 or Higher (OTL = CONTWR), LANG Fixed at 1.49

	CONTLWR level				
	1	2	3	4	5
Females					
EL	.20	.25	.30	.36	.42
Non EL	.16	.26	.39	.54	.68
Males					
EL	.13	.16	.20	.25	.30
Non EL	.10	.17	.27	.41	.56

On average, EL status had a negative effect (–.543) on performance. The effect of CONTWR on the EL effect was also negative (–.332). This indicates that, holding GENDER and LANG constant, EL students have lower scores than non-EL students on average, and the gap between EL and non-EL students gets larger as the class CONTWR level increases (see Figure 5).[4]

[4] There are various points of view in setting the *p*-value for α or Type 1 error. Type 1 error (α) is the error of rejecting the null hypothesis when the null hypothesis is true, and the *p*-value shows the probability of making type 1 error. Conventionally, α = 0.05 is used as a criterion of decision; however, α = 0.10 can be used as the *p*-value based on the sample size, complexity of the model, etc. Thus, setting the *p*-value at the 0.10 level, the differential impact of content coverage–writing is significant.

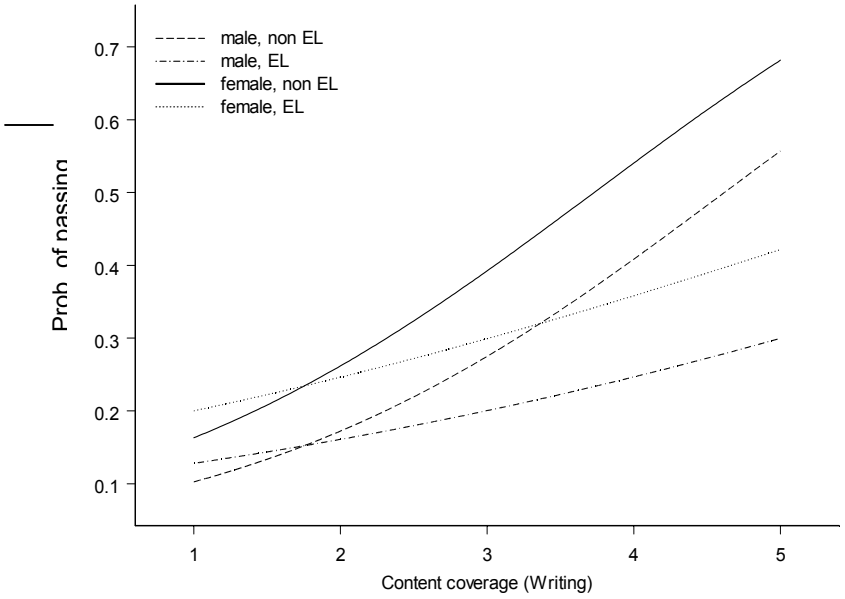


Figure 5. Differential impact of CONTWR on gender and language proficiency status.

Summary and Conclusions

The purpose of this initial study was to examine the factors that contribute to high performance of English Learners on standards-based performance assessments (e.g., the LAPA). Specifically, we were interested in identifying the Opportunity to Learn (OTL) variables that positively impact student performance. We also investigated potential differences in the impact of OTL on performance between ELs and non-ELs.

Our study suggests that there are several factors contributing to students' performance on the LAPA. At the student level, our analysis suggests that the greatest indicators of and contributors to individual students' LAPA scores were performance on the SAT-9 language test, ethnicity, gender, and language proficiency status. The highlights of the results are presented below.

- Female students on average performed better on the LAPA than male students.
- Performance on the SAT-9 Language test is related to LAPA performance. This information also serves as one source of evidence for construct validity of the LAPA.

- African American students performed lower on the LAPA compared to Hispanic students after controlling for language proficiency.
- ELs performed lower than their non-EL counterparts (i.e., IFEP, RFEP, and EO students) even after controlling for all other variables.

At the teacher level, we found that one OTL variable was statistically significantly associated with student performance. Consistent with results from previous studies utilizing this survey instrument (Boscardin et al., 2002), content coverage was a significant OTL factor. Our study showed that higher levels of content coverage in both writing and literary analyses were associated with higher performance for all students, including ELs. We also found differential impact of one OTL variable, content coverage–writing, on EL student performance. This finding indicates that the gap between ELs and non-ELs increases as teacher reports of content coverage–writing increase. In other words, the magnitude of the content coverage–writing is associated with the language proficiency level of students. The pattern for the relationship between content coverage–literary analysis and EL status was similar, though not significant.

This pattern of results may be reflective of teachers' inability to make the content accessible to ELs. Comprehensive literary analysis would involve highly abstract ideas about what the author is attempting to convey to the reader and a solid understanding of academic writing to convey those ideas in a written-response-to-literature task. In order to arrive at reasonable interpretations of a given text, ELs may need additional support from teachers due to the linguistic demands of the task. For example, constructing meaning from text involves not only knowledge about the words in the text, but also about the linguistic elements of the text including pragmatics of the English language, which are culturally bound (Kern, 2000). That is, students must have both cultural knowledge and knowledge of English conventions for expression, understanding, and interpretation to fully understand the author's message. Furthermore, students need to understand the patterns of academic written discourse to convey their interpretations of the text in writing. As pointed out earlier, in order to assist ELs in meeting the content and linguistic demands of the LAPA, teachers would be required to go beyond discussions with students of content to an analysis of the language used in the texts for rhetorical and aesthetic effect, in addition to an analysis of the language of a written character study. This aspect of academic language is often absent in English language arts instruction. In order to implement this type of instruction well,

teachers would need to have knowledge of structural linguistics and discourse analysis, which is not often incorporated into teacher training programs (Wong Fillmore & Snow, 2000). Such knowledge alone, however, may not yield the needed results. Teachers also need training on how to incorporate this content with strategies that have been shown to be effective for ELs, such as the Cognitive Academic Language Learning Approach (CALLA; Chamot & O'Malley, 1994).

The differences in the impact of OTL also suggest that there may be a unique set of experiences to which ELs need exposure before they can fully benefit from a standards-based reform effort utilizing performance assessments. For example, there is growing evidence that good readers arrive at school with more refined discourse skills (August & Hakuta, 1997). For ELs, first language oral and reading proficiency has been linked to higher English reading (Jimenez, Garcia, & Pearson, 1995; Lanauze & Snow, 1989; Moll & Diaz, 1985). Moreover, some researchers have found positive correlations between English second-language oral proficiency and English second-language reading ability (see Fitzgerald, 1995 for a review). Thus, oral language practice in the first and second languages is a potential EL directed instructional activity that can be incorporated into an OTL measure. If accountability systems are to provide accurate information about student achievement, it is necessary to identify such OTL dimensions and investigate the feasibility of utilizing survey instruments to capture these differences and their impact on student performance. The next phase of this project is intended to move precisely in this direction.

Implications and Next Steps

This study underscores the need to collect data to determine the impact of reform on achievement in general and for ELs in particular. Simply disaggregating the performance data to show potential disparities between groups of students does not provide policymakers and other stakeholders a complete picture of student achievement. Nor does such an approach provide information that leads to targeted areas for improvement. Narrow definitions of content coverage also may not yield the most information about the factors that contribute to student achievement. The definition of content coverage that was used in this initial study was limited to general topics related to the content of the performance assessment. A better strategy would be to include content knowledge items that are particularly relevant for ELs in this context, such as instruction in the academic language associated with

the task. Such a definition would lead to a better understanding of the factors that contributed to this differential impact.

Historically, economically disadvantaged and culturally diverse groups have had less access to challenging curriculum and other resources. The results of this initial study place an interesting spin on how this should be remedied, as it appears that providing greater amounts of content coverage alone—as the standards-based reform implies—does not lead to a decrease in the achievement gap between ELs and their non-EL counterparts. Therefore, future research on OTL should aim to identify OTL strategies that specifically address and target the special needs of the EL population. Such research may result in the development of an OTL indicator system that is more sensitive in investigating the impact of reform on EL achievement, in general, and more specifically in detecting the kinds of experiences ELs need that would prepare them to achieve expected standards.

The next set of studies of this project aims to investigate the extent to which instruction in academic language contributes to EL student achievement. The following are key operational goals of the next phase of this study.

- Observe and interview teachers whose EL students have demonstrated high achievement on the LAPA to develop teacher capacity-building strategies for increasing alignment of instruction and assessment and improve EL achievement.
- Test teacher capacity-building strategies to increase alignment of instruction and assessment and improve EL achievement.
- Identify factors that influence the implementation and consequences of standards-based assessments for ELs.
- Develop survey instruments designed to measure classroom factors that contribute to increased EL achievement.

The results of the second phase will be used to develop refined strategies for improving EL performance, and these strategies will be tested in subsequent phases of this project.

References

Abedi, J., Leon, S., & Mirocha, J. (2000). *Examining ELL and non-ELL student performance differences and their relationship to background factors: Continued analyses of extant data.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Academy Press.

Baker, E. L., Aschbacher, P., Niemi, D., & Sato, E. (1992). *CRESST performance assessment models: Assessing content area explanations.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E. L., Freeman, M., & Clayton, S. (1991). Cognitive assessments of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 131-153). Englewood Cliffs, NJ: Prentice-Hall.

Baker, E. L., Niemi, D., Herl, H., Aguirre-Muñoz, Z., Staley, L., & Linn, R. L. (1995). *Report on the Content Area Performance Assessment (CAPA): A collaboration among the Hawaii Department of Education, the Center for Research on Evaluation, Standards and Student Testing (CRESST), and the teachers and children of Hawaii*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bailey, A. L. (2000). *Language analysis of standardized achievement tests: Considerations in the assessment of English language learners.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bailey, A. L., & Butler, F. A. (2002). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Boscardin, C., Stoker, G., Kim, J., Kim, M., & Aguirre-Muñoz, Z. (2002). *An investigation into the validity of the Chicago Academic Standards Examinations: Year 3 report.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328-375). New York: Macmillan.

Chamot, A., & O'Malley, J. M. (1994). *The CALLA handbook. Implementing the Cognitive Academic Language Learning Approach.* Reading, MA: Addison-Wesley.

Cocking, R. R., & Chipman, S. (1988). Conceptual issues related to mathematics achievement in language minority children. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 17-46). Hillsdale, NJ: Erlbaum.

Collier, V. (1995). Acquiring a second language for school [Entire Issue]. *Directions in Language and Education, 1*(4).

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3-49). Los Angeles: California State University, Evaluation, Dissemination and Assessment Center.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy.* Clevedon, England: Multilingual Matters.

Darling-Hammond, L. (1990). Achieving our goals: Superficial or structural reforms? *Phi Delta Kappan, 72,* 286-295.

Darling-Hammond, L. (Ed.). (1994). *Professional development schools: Schools for developing a profession.* New York: Teachers College Press.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 7-32). Washington, DC: National Institute of Education.

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

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

Fitzgerald, J. (1995). English-as-a-second-language learners' cognitive reading processes: A review of research in the United States. *Review of Educational Research, 65,* 145-190.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms.* New York: Cambridge University Press.

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.

Jackson, K. W. (1982, March). *Achievement differences in mathematics: A search for causal explanations.* Paper presented at the annual meeting of the National Council of Black Studies, Chicago.

Jimenez, R. T., Garcia, G. E., & Pearson (1995). Three children, two languages, and strategic reading: Case studies in bilingual/monolingual reading. *American Educational Research Journal, 32,* 67-97.

Kern, R. (2000). *Literacy and language teaching.* Oxford: Oxford University Press.

Kozol, J. (2000). Forward. In J. Cohen & J. Rogers (Eds.), *Will standards save public education?* Boston: Beacon Press.

Lanauze, M., & Snow, C. E. (1989). The relation between first- and second-language writing skills: Evidence from Puerto Rican elementary school children in bilingual programs. *Linguistics and Education, 1,* 323-340.

Leinhardt, G. (1983). Overlap: Testing whether it is taught. In G. Madaus (Ed.), *The courts, validity, and minimum competency testing* (pp. 151-170). Boston: Kluwer-Nijhoff.

Leinhardt, G., Zigmond, N., & Cooley, W. (1981). Reading instruction and its effects. *American Educational Research Journal, 18,* 343-361.

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

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

McDonnell, L. M., Burstein, L., Ormseth, T., Catterall, J., & Moody, D. (1990). *Discovering what schools really teach: Designing improved coursework indicators* (Report No. R3886). Santa Monica, CA: RAND Corporation.

Moll, D. E., & Diaz, R. (1985). Ethnographic pedagogy: Promoting effective bilingual instruction. In E. Garcia & R. Padilla (Eds.), *Advances in bilingual education research* (pp. 127-149). Tucson: University of Arizona Press.

Muthén, B. O., Kao, C. F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement, 28,* 1-22.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

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

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

Saxe, G. B., Gearhart, M., & Seltzer, M. (1999). Relations between classroom practices and student learning in the domain of fractions. *Cognition and Instruction, 17,* 1-24.

Stevens, R. A., Butler, F. A., & Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of English language learners (ELLs)* (CSE Tech. Rep. No. 552). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Stevenson, H., & Stigler, J. (1992). *The learning gap: Why are schools are failing and what we can learn from Japanese and Chinese education.* New York: Summit Books.

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

Winfield, L. F. (1991). Resilience, schooling, and development in African-American youth: A conceptual framework. *Education and Urban Society, 24*(1), 5-14.

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

Wong Fillmore, L. (1991). Second language learning in children: A model of language learning in social context. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 49-69). Cambridge: Cambridge University Press.

Wong Fillmore, L., & Snow, C. E. (2000). *What teachers need to know about language. ERIC Clearinghouse on Languages and Linguistics Special Report.* Washington, DC: U.S. Department of Education Office of Educational Research and Improvement. (ERIC Document Reproduction Service Number ED 444379)

Appendix A:

Assessment Rubric

Rubric

The following rubric is used to assess student reading and writing skills:

ADVANCED — SCORE = 4

The response demonstrates well-developed reading comprehension skills and the ability to analyze a major literary element (characterization).

- Most of the important character features are described clearly and thoroughly. (Features may include physical and personality traits, thoughts and motivations, actions, relationships with other characters, and the character's impact on the story.)[1]
- Statements about the heroic qualities of a character are well supported or explained through references to the text.[2]
- Ideas are logically organized.
- Minor mechanical errors may be present but do not impede communication in most of the response.

PROFICIENT — SCORE = 3

The response demonstrates solid reading comprehension skills and the ability to analyze a major literary element (characterization).

- Some of the important character features are described clearly.
- Some statements about the heroic qualities of a character are generally supported or explained through references to the text.
- Most ideas are logically organized.
- Mechanical errors may be present but do not impede communication in most of the response.

PARTIALLY PROFICIENT — SCORE = 2

The response demonstrates some reading comprehension skills and the ability to analyze a major literary element (characterization).

- Few character features are described clearly, these features may not be heroic qualities.
- There is an attempt to use references to the text to support or explain the heroic qualities of a character.
- Some ideas are logically organized.
- Mechanical errors may impede communication in most of the response.

NOT PROFICIENT — SCORE = 1

The response demonstrates little or no skill in reading comprehension nor the ability to analyze a major literary element (characterization).

- Character features are not described, or the descriptions are unclear.
- Statements about the heroic qualities of a character are not supported or explained through references to the text.
- Ideas are not logically organized or are not provided.
- Many mechanical errors may impede communication throughout the response.

[1] This definition of character features applies to each score level.

[2] In general, sentences should not be copied directly from the text unless the student is using a quotation for a particular purpose.

Appendix B

Teacher Opportunity to Learn Survey

TEACHER SURVEY: GRADE 6

ID: _____

*Below are questions regarding your educational background, classroom practices, the [District] Language Arts Performance Assignment (ILAP), and classroom resources. Please answer the following questions frankly. Responses to these questions will be used to determine the factors that influence student achievement and for making recommendations to the District about this assessment program. Be assured that the answers you provide will only be reviewed and used for these purposes by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). This information will **not** be used for evaluation of teacher performance*

_____

Teacher Name: (Last, First, MI)

_____

Complete School Name

EDUCATIONAL BACKGROUND

1. What is the number of years that you have taught, including this year?
   a. Total teaching: _____ years
   b. At this school: _____ years
   c. This grade or course _____ years

2. Please indicate the approximate number of English or Language Arts content and/or methods courses you have taken at the following levels.

	None	1 – 3	4 – 6	7 – 9	10 or more
a. Undergraduate	O	O	O	O	O
b. Graduate	O	O	O	O	O

3. In what general field did you major as an undergraduate?
   O English/ Literature O Mathematics [or related] O Sciences O Humanities/History
   O Other

4. In what field is your master's degree?
   O English/ Literature O Mathematics [or related] O Sciences O Humanities/History
   O Education O Other
   O I do not have a master's degree O I am currently enrolled in a master's program

5. Do you have an advanced degree, such as an Ed.D. or a Ph. D.? O Yes ONo O Currently Enrolled in Program

6. What type of teaching credential do you possess? (Check all that apply):
   O Elementary O Single Subject Clear, subject _____
   O CLAD O Multi – Subject Clear, subjects _____
   O BCLAD O Emergency, type _____

7. Write the names (and topics) of recent (in the last 3 years) professional development seminars or workshops you have attended:

_____ _____ _____
_____ _____ _____
_____ _____ _____
_____ _____ _____

8. Have you been trained in ESL? O Yes ONo

9. Have you been trained in Sheltered Instruction/SDAIE? O Yes ONo

10. Please rate your level of expertise in each of the following:

	Novice	Adequate	Expert
a. Analyzing the plot (i.e., beginning, middle and end) of literary works	1	2	3
b. Analyzing actions in literary works	1	2	3
c. Using information for the literary works to support ideas or judgments about the story referencing literary works to support literary analysis	1	2	3

CLASSROOM PRACTICES

11. How much class time was spent learning about or doing each of the following in your class(es) during this school year? (Mark only one per item):

	less than 1 week	1 week to less than 2 weeks	2 weeks to less than 3 weeks	3 weeks to less than 4 weeks	4 or more weeks	*Don't remember*
a. Describing the plot or theme of novels, plays, or short stories	1	2	3	4	5	6
b. Describing heroic qualities of characters	1	2	3	4	5	6
c. Describing characters' physical or personality traits	1	2	3	4	5	6
d. Describing characters' motivations, thoughts, and feelings	1	2	3	4	5	6
e. Describing characters' actions or relationship with other characters	1	2	3	4	5	6
f. Using information from novels, plays, or short stories read to support ideas	1	2	3	4	5	6
g. Writing about heroic qualities of characters, sacrifices they make or how they are courageous	1	2	3	4	5	6
h. Writing about other aspects of characters, like physical traits, their relationship with other characters, or impact on the story	1	2	3	4	5	6

12. On average how often did you engage students in each of the following activities in your class(es) during this school year?

	Never or almost never	About once a month	About twice a month	Once or twice a week	Almost every day	*Don't remember*
a. Complete assignments or tests on grammar	1	2	3	4	5	6
b. Complete worksheets or exercises on spelling or vocabulary	1	2	3	4	5	6
c. Listen to the teacher discuss a topic for most of the class period	1	2	3	4	5	6
d. Read literary works (novels, short stories, poetry, essays or plays)	1	2	3	4	5	6
e. Use pre-writing activities (e.g., clustering, webbing, or brainstorming) to organize ideas	1	2	3	4	5	6
f. Write about literature discussed in class	1	2	3	4	5	6
g. Revise writing to clarify ideas, improve logic, organization, or spelling	1	2	3	4	5	6

13. How often did you use the following assessments strategies during this school year?

	Never or almost never	About once a month	About twice a month	Once or twice a week	Almost every day
a. Analyze students' writing assignments to assess their knowledge of literary elements	1	2	3	4	5
b. Use multiple-choice tests to assess students' knowledge and/or skills	1	2	3	4	5
c. Use constructed-response tests to assess students' knowledge and/or skills (e.g., open-ended, short answer)	1	2	3	4	5
d. Use other assessments to evaluate students' knowledge and/or skills (e.g., writing poems, research reports)	1	2	3	4	5
e. Use performance assessments to evaluate students' knowledge and/or skills (e.g., formal debate, presentations, demonstrations)	1	2	3	4	5
f. Use portfolios of student work to assess and monitor students' knowledge and/or skills	1	2	3	4	5
g. Assess understanding of key vocabulary and content concepts	1	2	3	4	5
h. Provide regular feedback to students on language and content work	1	2	3	4	5
i. Evaluate student understanding of forms and functions of English	1	2	3	4	5

14. How often do you incorporate the following sheltered instructional strategies into your lessons during the school year?

	Never	Less than once per week	Once per week	2-4 times per week	Once per day	2 or more times per day	Not Applicable*
a. Use supplementary materials (e.g., graphs, models, visuals) to make lessons clear and meaningful	1	2	3	4	5	6	7
b. Adapt content (e.g., text, assignments) to all levels of students' English proficiency	1	2	3	4	5	6	7
c. Explicitly link new concepts to students' background experiences and past learning	1	2	3	4	5	6	7
d. Use speech appropriate for students' English proficiency	1	2	3	4	5	6	7
e. Use scaffolding techniques to support students' understanding	1	2	3	4	5	6	7
f. Provide opportunities for student/teacher and student/student interactions that encourage elaborated responses	1	2	3	4	5	6	7
g. Provide activities for students to apply content and knowledge	1	2	3	4	5	6	7
h. Provide opportunities for students to clarify key concepts in primary language	1	2	3	4	5	6	7

15. How confident do you feel in the following areas?

	Not at all confident	Slightly confident	Somewhat confident	Very confident	Not Applicable*
a. Supporting English learners' (i.e., LEP students) access the ELA curriculum	1	2	3	4	5
b. Preparing English learners for the ILAP	1	2	3	4	5
c. Assessing English learners' English proficiency	1	2	3	4	5

**Only check "Not Applicable" if you do not have any English learners in your class(es).*

[DISTRICT] LANGUAGE ARTS PERFORMANCE ASSIGNMENT (ILAP)

16. How much class time was spent preparing for the ILAP doing the following activities?

	2 days or less	3 to 5 days	Between 1 and 2 weeks	2 weeks or more
a. Discussing and reviewing the ILAP rubric	O	O	O	O
b. Discussing the ILAP Anchor papers	O	O	O	O
c. Completing practice assignments similar to the ILAP	O	O	O	O
d. Reviewing important concepts necessary for completing the ILAP	O	O	O	O
e. Reviewing techniques for organizing ideas in written responses	O	O	O	O

CLASSROOM RESOURCES

17. Are any of the following available in your classroom?

	Yes	No
a. Dictionary available for most students	O	O
b. Thesaurus available for most students	O	O
c. Extensive classroom library	O	O
d. Computer(s) with literacy software	O	O

Appendix C

CFA Output

Appendix C: CFA Output

```
Mplus VERSION 2.01

MUTHEN & MUTHEN

12/02/2002  10:21 AM


INPUT INSTRUCTIONS


  TITLE: This is the MODIFIED CFA for grade 6

    DATA:

       File is tsg6.dat;

       Format is 592x, 34F10.0;


    VARIABLE:

       NAMES ARE yex1-yex3 ya1-ya8 yb1-yb7 yc1-yc6 yd1-yd5 ye yf1-yf4;

       USEVARIABLE ARE yex1-yex3 ya1-ya8 yb5-yb7 yc1-yc6 yd1-yd5 yf1-
yf4;

       CATEGORICAL ARE ALL;

       missing are all (9);

    MODEL:

         EXPERT  BY yex1-yex3;

         cntntlit BY ya1-ya5;

         cntntwr BY ya6-ya8;

         clsprcss BY yb5-yb7;

         ASSESS BY yc1-yc6;

         LAPAPREP BY yd1-yd5;

         RESOURCE BY yf1-yf4;
```

INPUT READING TERMINATED NORMALLY

This is the MODIFIED CFA for grade 6

SUMMARY OF ANALYSIS

Number of groups	1
Number of observations	28
Number of y-variables	29
Number of x-variables	0
Number of continuous latent variables	7

Observed variables in the analysis

YEX1	YEX2	YEX3	YA1	YA2	YA3
YA4	YA5	YA6	YA7	YA8	YB5
YB6	YB7	YC1	YC2	YC3	YC4
YC5	YC6	YD1	YD2	YD3	YD4
YD5	YF1	YF2	YF3	YF4	

Categorical variables

YEX1	YEX2	YEX3	YA1	YA2	YA3
YA4	YA5	YA6	YA7	YA8	YB5
YB6	YB7	YC1	YC2	YC3	YC4
YC5	YC6	YD1	YD2	YD3	YD4
YD5	YF1	YF2	YF3	YF4	

Continuous latent variables in the analysis

EXPERT	CNTNTLIT	CNTNTWR	CLSPRCSS	ASSESS	LAPAPREP
RESOURCE					

This is the MODIFIED CFA for grade 6

Estimator	WLSMV
Maximum number of iterations	1000
Convergence criterion	0.500D-04

Input data file(s)

tsg6.dat

Input data format

(592X, 34F10.0)

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	35.629*
Degrees of Freedom	19**
P-Value	0.0117

* The chi-square value for MLM, MLMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM chi-square difference testing is described on page 360 in the Mplus User's Guide.

** The degrees of freedom for MLMV and WLSMV are estimated according to formula 110 (page 358) in the Mplus User's Guide.

Chi-Square Test of Model Fit for the Baseline Model

Value	**********
Degrees of Freedom	17
P-Value	0.0000

CFI/TLI

CFI	1.000
TLI	1.000

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.177

SRMR (Standardized Root Mean Square Residual)

Value	0.187

WRMR (Weighted Root Mean Square Residual)

Value	1.008

This is the MODIFIED CFA for grade 6

MODEL RESULTS

	Estimates	S.E.	Est./S.E.

EXPERT BY			
YEX1	1.000	0.000	0.000
YEX2	1.028	0.025	40.427
YEX3	0.986	0.023	42.008
CNTNTLIT BY			
YA1	1.000	0.000	0.000
YA2	1.011	0.112	9.033
YA3	1.003	0.109	9.163
YA4	0.980	0.111	8.832
YA5	1.025	0.094	10.897
CNTNTWR BY			
YA6	1.000	0.000	0.000
YA7	0.824	0.112	7.335
YA8	0.691	0.101	6.860
CLSPRCSS BY			
YB5	1.000	0.000	0.000
YB6	1.272	0.169	7.542
YB7	1.375	0.188	7.314
ASSESS BY			
YC1	1.000	0.000	0.000
YC2	0.558	0.139	4.009
YC3	0.549	0.113	4.847
YC4	0.997	0.099	10.041

YC5	0.612	0.112	5.442
YC6	0.320	0.121	2.638
LAPAPREP BY			
YD1	1.000	0.000	0.000
YD2	1.037	0.178	5.813
YD3	1.140	0.186	6.120
YD4	1.343	0.143	9.385
YD5	1.295	0.152	8.517
RESOURCE BY			
YF1	1.000	0.000	0.000
YF2	0.886	0.282	3.138
YF3	0.512	0.226	2.263
YF4	0.861	0.315	2.733
CNTNTLIT WITH			
EXPERT	0.300	0.107	2.811
CNTNTWR WITH			
EXPERT	0.106	0.133	0.799
CNTNTLIT	0.704	0.079	8.877
CLSPRCSS WITH			
EXPERT	0.269	0.093	2.879

This is the MODIFIED CFA for grade 6

CNTNTLIT	0.364	0.113	3.220
CNTNTWR	0.513	0.084	6.138

ASSESS WITH

EXPERT	0.364	0.119	3.050
CNTNTLIT	0.387	0.088	4.423
CNTNTWR	0.566	0.073	7.741
CLSPRCSS	0.625	0.094	6.615

LAPAPREP WITH

EXPERT	0.065	0.107	0.603
CNTNTLIT	0.166	0.079	2.111
CNTNTWR	0.521	0.075	6.957
CLSPRCSS	0.398	0.081	4.907
ASSESS	0.462	0.065	7.102

RESOURCE WITH

EXPERT	0.701	0.205	3.422
CNTNTLIT	0.193	0.133	1.448
CNTNTWR	0.103	0.198	0.523
CLSPRCSS	0.421	0.191	2.204
ASSESS	0.526	0.227	2.319
LAPAPREP	0.288	0.142	2.021

Variances

EXPERT	0.973	0.024	40.428
CNTNTLIT	0.793	0.138	5.741
CNTNTWR	1.035	0.159	6.492
CLSPRCSS	0.477	0.128	3.720
ASSESS	0.751	0.103	7.300
LAPAPREP	0.559	0.122	4.602
RESOURCE	0.971	0.603	1.611

Beginning Time: 10:21:20

Ending Time: 10:21:24

Elapsed Time: 00:00:04

MUTHEN & MUTHEN

11965 Venice Blvd., Suite 407

Los Angeles, CA 90066

Appendix D

Ordinal Logistic HLM Analysis Output

APPENDIX D:

ORDINAL LOGISTIC HLM ANALYSIS OUTPUT

```
 Summary of the model specified (in equation format)

 ----------------------------------------------------


Level-1 Model


  Prob[R = 1|B] = P'(1) = P(1)

  Prob[R <= 2|B] = P'(2) = P(1) + P(2) + P(3)

  Prob[R <= 3|B] = P'(3) = P(1) + P(2) + P(3)

  Prob[R <= 4|B] = 1.0


  where


  P(1) = Prob[Y(1) = 1|B]

  P(2) = Prob[Y(2) = 1|B]

  P(3) = Prob[Y(3) = 1|B]


  log[P'(1)/(1 - P'(1)] = B0 + B1*(FEMALE) + B2*(SCLANGPR) + B3*(ELL)

  log[P'(2)/(1 - P'(2)] = B0 + B1*(FEMALE) + B2*(SCLANGPR) + B3*(ELL) + d(2)

  log[P'(3)/(1 - P'(3)] = B0 + B1*(FEMALE) + B2*(SCLANGPR) + B3*(ELL) + d(3)



Level-2 Model

  B0 = G00 + G01*(CONTLIT) + U0
```

```
 B1 = G10

 B2 = G20

 B3 = G30 + G31*(CONTLIT)

RESULTS FOR ORDINAL ITERATION 9




 Tau

 INTRCPT1,B0      0.69703




Tau (as correlations)

 INTRCPT1,B0  1.000


 ----------------------------------------------------

  Random level-1 coefficient   Reliability estimate

 ----------------------------------------------------

  INTRCPT1, B0                        0.780

 ----------------------------------------------------


The value of the likelihood function at iteration 2 = -1.847082E+003
```

The outcome variable is LSCORER

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1 slope, B0					
INTRCPT2, G00	-2.602236	0.226064	-11.511	21	0.000
CONTLIT, G01	0.464062	0.194889	2.381	21	0.027
For FEMALE slope, B1					
INTRCPT2, G10	0.529681	0.140571	3.768	791	0.000
For SCLANGPR slope, B2					
INTRCPT2, G20	0.480522	0.032647	14.719	791	0.000
For ELL slope, B3					
INTRCPT2, G30	-0.540861	0.194455	-2.781	791	0.006
CONTLIT, G31	-0.218281	0.161746	-1.350	791	0.177
For THOLD2,					
d(2)	2.026945	0.124003	16.346	791	0.000
For THOLD3,					
d(3)	4.673779	0.183373	25.488	791	0.000

The outcome variable is LSCORER

Final estimation of fixed effects

(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1 slope, B0					
INTRCPT2, G00	-2.602236	0.252989	-10.286	21	0.000
CONTLIT, G01	0.464062	0.192493	2.411	21	0.025
For FEMALE slope, B1					
INTRCPT2, G10	0.529681	0.161768	3.274	791	0.001
For SCLANGPR slope, B2					
INTRCPT2, G20	0.480522	0.029736	16.159	791	0.000
For ELL slope, B3					
INTRCPT2, G30	-0.540861	0.151550	-3.569	791	0.001
CONTLIT, G31	-0.218281	0.121053	-1.803	791	0.071
For THOLD2,					
d(2)	2.026945	0.197793	10.248	791	0.000
For THOLD3,					
d(3)	4.673779	0.273372	17.097	791	0.000

The robust standard errors are appropriate for datasets having a moderate to large number of level 2 units. These data do not meet this criterion.

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	0.83488	0.69703	21	142.64106	0.000

Summary of the model specified (in equation format)

Level-1 Model

Prob[R = 1|B] = P'(1) = P(1)

Prob[R <= 2|B] = P'(2) = P(1) + P(2) + P(3)

Prob[R <= 3|B] = P'(3) = P(1) + P(2) + P(3)

Prob[R <= 4|B] = 1.0

where

P(1) = Prob[Y(1) = 1|B]

P(2) = Prob[Y(2) = 1|B]

P(3) = Prob[Y(3) = 1|B]

log[P'(1)/(1 - P'(1)] = B0 + B1*(FEMALE) + B2*(SCLANGPR) + B3*(ELL)

log[P'(2)/(1 - P'(2)] = B0 + B1*(FEMALE) + B2*(SCLANGPR) + B3*(ELL) + d(2)

log[P'(3)/(1 - P'(3)] = B0 + B1*(FEMALE) + B2*(SCLANGPR) + B3*(ELL) + d(3)

Level-2 Model

B0 = G00 + G01*(CONTWR) + U0

B1 = G10

B2 = G20

B3 = G30 + G31*(CONTWR)

RESULTS FOR ORDINAL ITERATION 8

Tau

INTRCPT1,B0 0.74054

Tau (as correlations)

INTRCPT1,B0 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.788

The value of the likelihood function at iteration 2 = -1.859014E+003

The outcome variable is LSCORER

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1 slope, B0					
INTRCPT2, G00	-2.579118	0.229204	-11.253	21	0.000
CONTWR, G01	0.542807	0.222748	2.437	21	0.024
For FEMALE slope, B1					
INTRCPT2, G10	0.532623	0.140513	3.791	791	0.000
For SCLANGPR slope, B2					
INTRCPT2, G20	0.485793	0.032696	14.858	791	0.000
For ELL slope, B3					
INTRCPT2, G30	-0.542779	0.194707	-2.788	791	0.006
CONTWR, G31	-0.332052	0.193793	-1.713	791	0.086
For THOLD2,					
d(2)	2.032521	0.124186	16.367	791	0.000
For THOLD3,					
d(3)	4.685334	0.183838	25.486	791	0.000

The outcome variable is LSCORER

Final estimation of fixed effects

(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1 slope, B0					
INTRCPT2, G00	-2.579118	0.225737	-11.425	21	0.000
CONTWR, G01	0.542807	0.179768	3.019	21	0.007
For FEMALE slope, B1					
INTRCPT2, G10	0.532623	0.161470	3.299	791	0.001
For SCLANGPR slope, B2					
INTRCPT2, G20	0.485793	0.029716	16.348	791	0.000
For ELL slope, B3					
INTRCPT2, G30	-0.542779	0.133438	-4.068	791	0.000
CONTWR, G31	-0.332052	0.145607	-2.280	791	0.023
For THOLD2,					
d(2)	2.032521	0.191667	10.604	791	0.000
For THOLD3,					
d(3)	4.685334	0.267880	17.490	791	0.000

The robust standard errors are appropriate for datasets having a moderate to large number of level 2 units. These data do not meet this criterion.

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	0.86055	0.74054	21	161.16721	0.000