ISSUES&ANSWERS REL 2009–No. 066

REL NORTHEAST & ISLANDS
Regional Educational Laboratory At Education Development Center, Inc.

New measures of English language proficiency and their relationship to performance on large-scale content assessments

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

ISSUES & ANSWERS REL 2009–No. 066

REL NORTHEAST & ISLANDS
Regional Educational Laboratory
At Education Development
Center, Inc.

New measures of English language proficiency and their relationship to performance on large-scale content assessments

January 2009

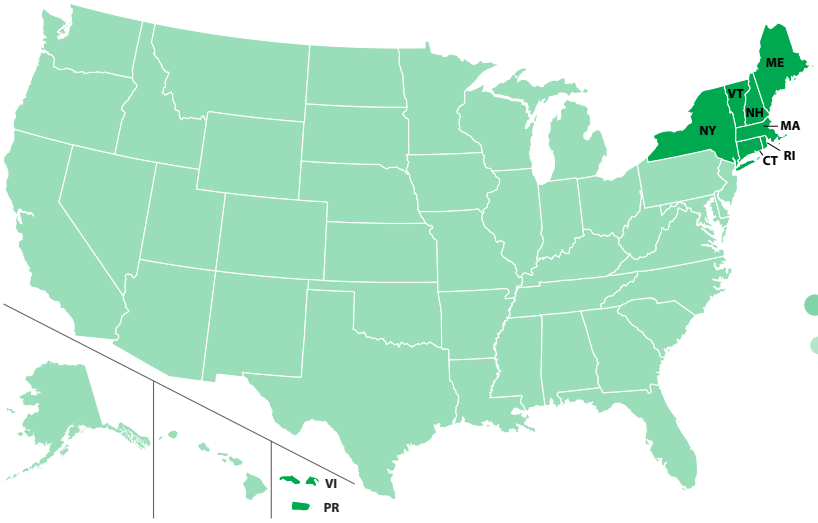
Prepared by

Caroline E. Parker
Education Development Center, Inc.

Josephine Louie
Education Development Center, Inc.

Laura O'Dwyer
Boston College

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education



REL
NORTHEAST
& ISLANDS
Regional Educational Laboratory
At Education Development
Center, Inc.

Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

January 2009

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0025 by Regional Educational Laboratory Northeast and Islands administered by Education Development Center, Inc. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Parker, C. E., Louie, J., and O'Dwyer, L. (2009). *New measures of English language proficiency and their relationship to performance on large-scale content assessments* (Issues & Answers Report, REL 2009–No. 066). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the regional educational laboratory web site at http://ies.ed.gov/ncee/edlabs.

Summary

REL 2009–No. 066

New measures of English language proficiency and their relationship to performance on large-scale content assessments

Using assessment results for 5th and 8th grade English language learner students in three Northeast and Island Region states, the report finds that the English language domains of reading and writing (as measured by a proficiency assessment) are significant predictors of performance on reading, writing, and mathematics assessments and that the domains of reading and writing (literacy skills) are more closely associated with performance than are the English language domains of speaking and listening (oral skills).

As the English language learner population grows throughout the Northeast and Islands Region, state departments of education are seeking assistance in creating comprehensive approaches to meeting English language learner students' academic needs in both instruction and assessment. Driving educator concerns is the fact that English language learner students consistently score lower on state assessments than students for whom English is their first language. In the context of the No Child Left Behind Act of 2001 (NCLB), states are seeking information to inform their efforts to reduce achievement gaps and to bring English language learner students, along with other traditionally underserved student subgroups, to proficiency on statewide assessments.

In response to a request from New Hampshire, Rhode Island, and Vermont to explore how English language proficiency measures may be related to performance outcomes on content assessments, this report uses the results of two new large-scale assessments—the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) English proficiency assessment and the New England Common Assessment Program (NECAP)—to address the following research question:

> *How does performance in four language domains on an English language proficiency assessment predict English language learner students' performance on a state content assessment after accounting for student and school characteristics?*

Based on findings from previous research, this report hypothesized that after controlling for individual student characteristics such as gender, poverty status, disability status, race/ethnicity, age for grade, and years in English language learner programs as well as

for school characteristics such as school size, school poverty, racial composition, English language learner student density, and geography, measures of academic English language proficiency would predict English language learner student outcomes on state content assessments. The report also hypothesized that measures of English language literacy (reading and writing) would be stronger predictors of content assessment outcomes than would measures of English oral proficiency (listening and speaking).[1]

To test these hypotheses, multilevel regression models were fit to assessment score data for 5th and 8th grade English language learner students in New Hampshire, Rhode Island, and Vermont. After controlling for student and school characteristics, English language proficiency scores (as measured by ACCESS) were indeed significant predictors of content assessment outcomes (as measured by the NECAP). The models also showed that after accounting for other covariates, ACCESS measures of English literacy were significantly stronger predictors of NECAP outcomes than were ACCESS measures of oral proficiency. Specifically, this report finds that:

- NECAP reading scores in both 5th and 8th grades were significantly and positively predicted by ACCESS reading, writing, and speaking scores after controlling for other ACCESS scores and student and school characteristics. Among the ACCESS domain scores the strongest predictor of NECAP reading outcomes was ACCESS reading scores, followed by ACCESS writing and speaking scores. ACCESS domain scores explained 30 percent of the variance in NECAP reading scores in 5th grade and 23 percent in 8th grade after controlling for student and school covariates.

- NECAP writing scores in 5th grade were significantly and positively predicted by ACCESS reading and writing scores and in 8th grade by all four ACCESS domain scores after controlling for other ACCESS scores and student and school characteristics. ACCESS reading and writing scores were the strongest predictors of NECAP writing outcomes in 5th and 8th grades. ACCESS domain scores explained 28 percent of the variance in NECAP writing scores in 5th grade and 25 percent in 8th grade after controlling for other covariates.

- Like NECAP reading and writing scores, NECAP mathematics scores in both 5th and 8th grades were positively and significantly predicted by ACCESS reading and writing scores after controlling for other ACCESS scores and student and school characteristics. Among the ACCESS domain scores ACCESS reading scores were the strongest predictor of NECAP mathematics outcomes for both 5th and 8th grade English language learner students, followed by ACCESS writing scores. ACCESS domain scores explained 21 percent of the variance in NECAP mathematics scores in 5th grade and 14 percent in 8th grade.

- ACCESS reading and writing scores were significant predictors of NECAP reading, writing, and mathematics scores in 5th and 8th grades. ACCESS speaking and listening scores were significant predictors of NECAP scores for only four outcomes: 5th and 8th grade reading (speaking), 8th

grade writing (speaking and listening), and 5th grade mathematics (listening).

In sum, ACCESS measures of English literacy skills (reading and writing scores) were significant predictors of NECAP reading and writing outcomes in 5th and 8th grades. Notably, ACCESS reading and writing scores were also positive and significant predictors of NECAP mathematics scores. In addition, except for 8th grade writing, ACCESS reading and writing scores were significantly stronger predictors of NECAP outcomes than were ACCESS listening and speaking scores. This evidence supports the original hypothesis that ACCESS measures of English literacy skills are better predictors of NECAP content outcomes than are ACCESS measures of English oral skills (listening and speaking). Readers are cautioned, however, that the analyses and interpretations presented are correlational and therefore do not allow causal conclusions.

In 5th and 8th grades, ACCESS scores explained 14–30 percent of the variance in scores for all three NECAP content scores (reading, writing, and mathematics) after controlling for background student and school characteristics. The ACCESS scores explained more of the variance in 5th grade (from 21 percent of NECAP mathematics scores to 30 percent of NECAP reading scores) than in 8th grade (from 14 percent of NECAP mathematics scores to 25 percent of NECAP writing scores).

January 2009

Note

1. In this report "stronger" predictors are defined as those whose regression coefficients are larger than those of other noted predictors in the study's regression models. A predictor is "significantly stronger" than another predictor when the difference between the regression coefficients is greater than zero at the $p < 0.05$ level.

TABLE OF CONTENTS

Using assessment results for 5th and 8th grade English language learner students in three Northeast and Island Region states, the report finds that the English language domains of reading and writing (as measured by a proficiency assessment) are significant predictors of performance on reading, writing, and mathematics assessments and that the domains of reading and writing (literacy skills) are more closely associated with performance than are the English language domains of speaking and listening (oral skills).

WHY THIS STUDY?

As the English language learner population grows throughout the Northeast and Islands Region, and as achievement gaps persist between English language learner students and native English speakers, state education agencies are creating comprehensive programs to meet English language learner student needs. With more than one in five school-age children in Rhode Island speaking a language other than English at home (Kids Count Data Center 2006), the Rhode Island Department of Education and the Governor's PK–16 Council have identified educating English language learner students as a priority. And in New Hampshire and Vermont, where English language learner populations are smaller and more isolated, state education agencies are looking for efficient ways to meet these students' needs. New Hampshire, for example, has recently requested assistance from regional education support centers to define and monitor services for English language learner students.

Regional need

In the context of the No Child Left Behind Act of 2001 (NCLB), Northeast and Islands Region states want technical assistance and targeted data analysis to inform their efforts to reduce achievement gaps and to bring English language learner students, along with members of other traditionally underserved student subgroups, to proficiency on statewide assessments. English language learner students consistently score lower on state assessments than native English speakers, often by as many as 20–30 percentage points (Abedi and Dietel 2004). The reasons for such low performance are varied and complex, not least of which is that English language learner students are learning content (mathematics, science, reading, and writing) and are being assessed in these content areas while they are learning the academic English that is the medium for classroom learning (see box 1 for a definition of key terms).

To better understand the learning needs of English language learner students, New Hampshire, Rhode Island, and Vermont have been administering a new

BOX 1

Definitions of key terms

Academic English. Researchers distinguish between social English and the academic English needed to learn academic content. Academic language uses different vocabularies, types of syntax, and levels of classroom discourse and involves abstract forms of language needed to communicate in formal, often decontextualized, situations and may be needed for successful navigation of classroom learning and large-scale assessments. (For more detail on the literature, see appendix A.)

English language learner student. Although definitions vary, the Council of Chief State School Officers defines an English language learner student as a student with a language background other than English and whose proficiency in English is such that the probability of the student's academic success in an English-only classroom is below that of an academically successful peer with an English language background (Council of Chief State School Officers 1992).

English language proficiency. Although definitions vary, the Council of Chief State School Officers defines a fully English proficient student as a student who is able to use English to ask questions, to understand teachers and reading materials, to test ideas, and to challenge what is being asked in the classroom. Four language skills contribute to proficiency:

- *Reading.* The ability to comprehend and interpret text at the age and grade-appropriate level.
- *Listening.* The ability to understand the language of the teacher and instruction, comprehend and extract information, and follow the instructional discourse through which teachers provide information.
- *Writing.* The ability to produce written text with content and format fulfilling classroom assignments at age- and grade-appropriate levels.
- *Speaking.* The ability to use oral language appropriately and effectively in learning activities (such as peer tutoring, collaborative learning activities, and question and answer sessions) within the classroom and in social interactions within the school (Council of Chief State School Officers 1992).

Multilevel regression modeling. A set of regression-based procedures used to analyze data with a nested or hierarchical structure (such as students nested within schools). When used with nested data, multilevel regression modeling allows correct standard errors to be calculated, allows the relationship between the independent and dependent variables to vary across groups, and allows individual and group characteristics to be included in models for predicting individual outcomes (Raudenbush and Bryk 2002).

Reliability estimate. Reliability is the consistency of measurement. A reliability estimate is a number calculated to represent the consistency of scores provided by a measurement instrument. The reliability estimates referred to here are internal consistency estimates of reliability (Cronbach's α). Calculating internal consistency reliability estimates requires only one administration of the measurement tool and is calculated from the interitem correlations. Values range from 0 to 1. Estimates of 0.7 or higher indicate optimal reliability.

Scale score. A scale score is a test score that has been converted from a raw score (such as a number correct) to a number on a common scale indicating a student's performance. NECAP scale scores range from 500 to 580 for grade 5 and from 800 to 880 for grade 8 in all content areas. ACCESS scale scores range from 100 to 600.

Standard deviation. Standard deviation is a measure of how widely or narrowly data are dispersed around the mean for the distribution. For example, the standard deviation of a set of student test scores is calculated by summing the squared deviations of each student's individual score from the mean, dividing this sum by one minus the total number of students, and taking the square root of the resulting number. A student's test score can be described in terms of standard deviation units by subtracting the mean from the student's score and dividing that figure by the standard deviation.

Standard error. Standard error is a measure of the amount of error between an estimated statistic from a sample and the true statistic for the population. For example, the mean test score for a sample of students will have a standard error that estimates the deviation between the sample mean and the mean for the entire student population. The standard error for a sample mean is calculated by dividing the standard deviation of the sample data by the square root of the number of subjects in the sample.

Variance. Variance is the standard deviation squared.

English language proficiency assessment called the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) since 2005. (Appendix A reviews previous and current generations of English language proficiency assessments, discusses the role that student and school characteristics may play in English language acquisition and performance on content assessments, and briefly surveys the literature on the relationship between English language proficiency and demonstration of context knowledge among English language learner students.) Unlike previous generations of English language proficiency assessments, ACCESS measures social and academic English in the four language domains of reading, writing, listening, and speaking. In addition to using the same English language proficiency assessment, the three states collaboratively defined grade-level expectations for all students and designed a common assessment for their state accountability systems, the New England Common Assessment Program (NECAP). Since 2005 the states have administered NECAP to assess student proficiency in reading, writing, and mathematics.[1]

With these new data on English language proficiency and content knowledge of English language learner students, New Hampshire, Rhode Island, and Vermont have requested assistance from REL Northeast and Islands to examine ACCESS and NECAP results, specifically to explore how English language proficiency measures may predict performance outcomes on content assessments. As a first step, the three states have jointly requested an examination of their combined data, hoping that the results will offer educators insight into the English language skills most highly correlated with better performance on content assessments.

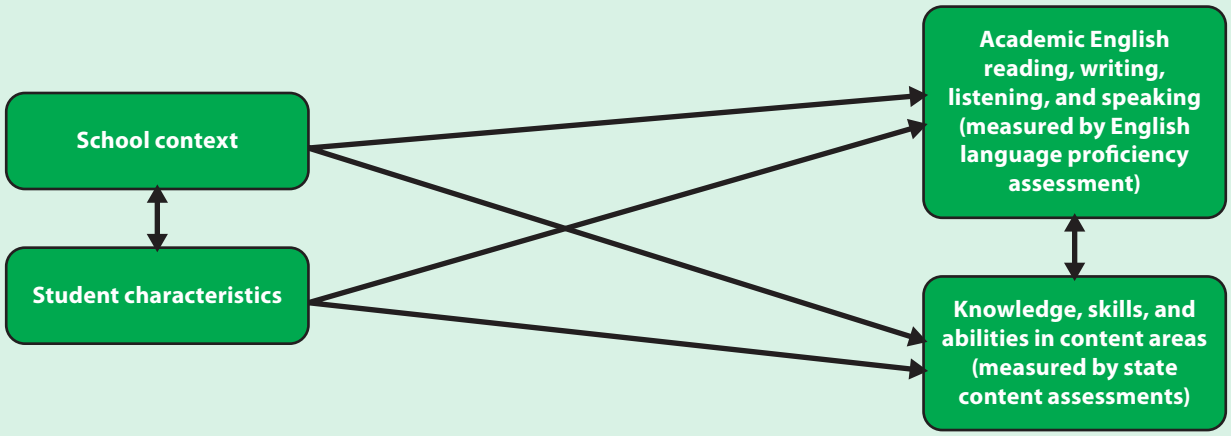
Research question and conceptual framework

This report set out to explore the following research question:

> *How does performance in four language domains on an English language proficiency assessment predict English language learner students' performance on a state content assessment after accounting for student and school characteristics?*

Figure 1 proposes a conceptual framework with sets of factors that may contribute to English language learner students' success in learning academic content, as demonstrated by scores on state content assessments. The framework outlines possible relationships among academic English skills within the four English language domains and school and student characteristics, which may influence both the English language acquisition process and the ability to demonstrate content knowledge. In turn, familiarity with academic

FIGURE 1
Conceptual framework: acquiring language of instruction and demonstrating knowledge, skills, and abilities on content assessment



Source: Authors' construction.

English may also directly affect ability to demonstrate content knowledge.

This study hypothesized that English language proficiency skills would significantly predict content assessment scores after controlling for student and school background variables. Due to the heavy use of reading and writing activities on large-scale assessments, this report also hypothesized that English skills in reading and writing (literacy skills) would be stronger predictors of performance on large-scale assessments than would English skills in listening and speaking (oral skills). The purpose of this report is to find preliminary evidence for these proposed relationships.

To answer the report's research question, researchers examined how well English language proficiency scores in four language domains (listening, speaking, reading, and writing) predict performance on content assessments in three areas (reading, writing, and mathematics) by English language learner students in New Hampshire, Rhode Island, and Vermont after controlling for other covariates (box 2 and appendix B discuss the report's methodology, and appendix C discusses the data used for the report). Measures of English language proficiency were from the 2006 ACCESS for ELLs English language proficiency assessment, which has been administered in the three states since 2006. Measures of content knowledge among

BOX 2

Methodology

The student covariates included in the analysis were gender, race/ethnicity, poverty status, disability status, age status (whether the student was overage for grade), and years in English language learner programs. The school covariates were student population size, school poverty, school racial composition, population density of English language learner students within the school (which this report refers to as "English language learner student density"), and school geographic location. To aid in the interpretation of results, all continuous covariates (years in English language learner programs and all the school variables) were grand-mean centered, and some (school size, school poverty, racial/ethnic composition, and English language learner student density) were rescaled (see appendix C).

Multilevel regression models were fit to the 5th and 8th grade English language learner student data to predict NECAP outcome variables using ACCESS scores and the student and school covariates described. Because regression analysis and observational data were used, the estimated relationships represent partial correlations and do not imply causation. Rather, the regression coefficients in the models describe the association between a dependent variable (for example, one of the NECAP scores) and the independent variables (ACCESS scores) while holding all other covariates (student and school characteristics) in the model constant.

Multilevel regression models were used to account for the interdependence of assessment scores among English language learner students attending the same schools. The percent of variation in NECAP outcome scores between schools (the intraclass correlation coefficient) was significant in all content areas and at both grades. In 5th grade the intraclass coefficient was 15.5 percent in reading, 20.5 percent in writing, and 13.0 percent in mathematics. In 8th grade it was 26.8 percent in reading, 28.2 percent in writing, and 16.4 percent in mathematics.

The multilevel regression models were fit to the 5th and 8th grade English language learner student data in stages. First, NECAP reading, writing, and mathematics scores were regressed on the student and school covariates (models 1 and 2). Then ACCESS scores were added to the model (model 3). Thus, in addition to the unconditional model that included only a random school effect, three models were fit to the 5th and 8th grade data samples for each NECAP outcome variable (reading, writing, and mathematics scores). Model 3 allowed the researchers to address the primary research question for this report.

Appendix F provides additional details on the multilevel modeling procedures used and explains the calculation of the percentage of variance. Appendix G presents the results of the multilevel models in which NECAP scores in reading, writing, and mathematics are regressed on the student and school covariates and the ACCESS domain scores for both the 5th and 8th grade samples.

these same students were from the 2006 NECAP, which has been administered in the three states since 2005. As noted, both ACCESS and NECAP are new research-based assessments that have been designed to maximize the reliability of student performance outcomes (see appendix D).

Data were examined specifically for English language learner students who took the 4th and 7th grade ACCESS assessments in spring 2006 and the 5th and 8th grade NECAP assessments in fall 2006.[2] The report focused on students who had taken the 5th and 8th grade NECAP because assessments in all three content areas (reading, writing, and mathematics) were administered to students in those two grades only.[3] Using multilevel regression modeling techniques, statistical relationships between English language learner student scores on the ACCESS and NECAP assessments were examined while controlling for student and school characteristics. Appendix C provides further detail on the two datasets (subsequently called the 5th and 8th grade English language learner samples) assembled for this report.

To compare ACCESS scores in the four English language domains and their relationships with NECAP reading, writing, and mathematics scores, all ACCESS and NECAP variables were standardized before they were incorporated into the multilevel models. Appendix C describes the variables, their original scales, and how they were recoded and rescaled prior to inclusion in the multilevel regression models.

HOW DOES PERFORMANCE IN FOUR LANGUAGE DOMAINS ON AN ENGLISH LANGUAGE PROFICIENCY ASSESSMENT PREDICT ENGLISH LANGUAGE LEARNER STUDENTS' PERFORMANCE ON A STATE CONTENT ASSESSMENT AFTER ACCOUNTING FOR STUDENT AND SCHOOL CHARACTERISTICS?

Table 1 presents predicted NECAP scores and changes in scores from the final multilevel models (model 3). The standardized regression coefficients are shown on the left in panel 1, and estimates in the original scale score points are shown on the right in panel 2. The standardized regression coefficients show which ACCESS domain scores are the strongest predictors of NECAP content scores holding other variables constant. For example, the standardized coefficients show that compared with ACCESS writing scores, ACCESS reading scores are stronger predictors of NECAP reading outcomes in 5th grade. The original scale score points show how well ACCESS scores predict NECAP scores in the original metrics of each assessment. Although the variability of the regression coefficients across ACCESS domains could be due to measurement error in the assessment scores, data indicate that NECAP and ACCESS provide reliable estimates of students' ability in the domains assessed (see appendix D).

Data were examined for English language learner students who took the 4th and 7th grade ACCESS assessments in spring 2006 and the 5th and 8th grade NECAP assessments in fall 2006

The top row of table 1 shows the intercept for each regression model, which is the predicted NECAP scale score when all predictors in the model are equal to the grand mean or zero (depending on how the variables were coded or centered).[4] The predicted NECAP scale scores at the intercept for the 5th grade sample were approximately 539 in reading, 534 in writing, and 540 in mathematics; for the 8th grade sample they were 833 in reading, 831 in writing, and 831 in mathematics.

Another indicator of how well ACCESS reading, writing, speaking, and listening scores predict NECAP reading outcomes is the percent of variance in NECAP reading scores that is explained by ACCESS domain scores within the model. Table 2 shows the variance explained in each of the models: model 1 includes only student covariates, model 2 includes student and school covariates, and model 3 includes, student and school covariates and ACCESS scores.

TABLE 1

NECAP scores regressed on different student ACCESS scores and student and school characteristics, 2006

	Panel 1: standard deviation units					
	Reading		Writing		Math	
Predictor	5th grade	8th grade	5th grade	8th grade	5th grade	8th grade
Intercept[a]	0.271***	0.234**	0.171	0.289**	0.373****	0.137
ACCESS predictors[b]						
Listening	0.031	0.018	0.048	0.110***	0.105***	0.001
Speaking	0.093****	0.097***	0.015	0.135****	0.028	0.015
Reading	0.383****	0.283****	0.303****	0.187****	0.320****	0.260****
Writing	0.239****	0.261****	0.294****	0.268****	0.147****	0.223****
Student characteristics						
Gender[c]	0.020	0.063	0.175****	0.199****	−0.132***	−0.110**
Poverty status[d]	−0.146****	−0.036	−0.074	0.000	−0.109	0.019
Disability status[e]	−0.337****	−0.487****	−0.312****	−0.337****	−0.384****	−0.497****
Age status[f]	−0.079	−0.100	−0.144	−0.046	−0.015	−0.024
Asian	−0.032	0.077	0.009	−0.076	−0.026	0.228**
Non-Hispanic Black	−0.298***	−0.019	−0.163	−0.103	−0.395****	−0.115
Hispanic	−0.130	−0.134	−0.100	−0.272***	−0.275****	−0.182**
Years in English language learner programs	−0.061****	−0.007	−0.047***	−0.006	−0.072****	−0.005
School characteristics						
School size[g]	0.013	−0.003	0.010	−0.010	0.001	−0.007
School poverty[h]	−0.005	−0.051	0.013	−0.014	−0.050**	−0.115***
Racial composition[i]	−0.006	−0.042	0.020	−0.046	−0.078***	−0.087***
English language learner student density[j]	0.017	−0.034	−0.015	−0.041	−0.004	−0.014
Rural	0.221**	0.161	0.112	−0.003	0.109	0.197
Urban	−0.024	−0.087	−0.038	−0.169	−0.041	0.033
New Hampshire	−0.066	−0.114	−0.160	−0.054	0.174	0.055
Vermont	0.011	0.114	−0.175	0.213	0.229	0.240

** Statistically significant at $p < 0.05$. *** Statistically significant at $p < 0.01$. **** Statistically significant at $p < 0.001$.

a. The predicted NECAP score for the English language learner student in the 5th or 8th grade sample who achieved the average score for the entire sample in each ACCESS domain and who was male, white, not living in poverty, not with disabilities, who spent an average number of years in English language learner programs, and who attended a Rhode Island suburban school of average size, poverty level, percent white, and English language learner density.

b. Predicted NECAP score changes measured in standard deviation units are from a 1 standard deviation unit increase in ACCESS scores, and predicted NECAP score changes measured in scale score points are from a 10 point increase in ACCESS scale scores. Readers are cautioned not to compare predicted changes in NECAP outcomes associated with 10 scale score point shifts in ACCESS domain scores. Ten-point score shifts are not equivalent across the four ACCESS domains because scores from each domain have different standard deviations. For example, 10 scale score points represent over a third of a standard deviation in 5th grade ACCESS writing scores but a seventh of a standard deviation in 5th grade ACCESS speaking scores (see table C6 in appendix C). Regression coefficients were calculated for 10 point shifts for all ACCESS predictors not to suggest that these shifts are equivalent but simply to facilitate the presentation of findings.

c. Defined as 1 = female, 0 = male.

d. Defined as 1 = in poverty (eligible for free or reduced-price lunch), 0 = not in poverty.

e. Defined as 1 = with disabilities (has an Individualized Education Program), 0 = without disabilities.

f. Age status was 1 = overage, 0 = not overage.

g. Measured in units of 100 students.

h. Defined as the percent of students within the school eligible for free or reduced-price lunch. Measured in units of 10 percentage points.

i. Defined as the percent of students in the school who are White. Measured in units of 10 percentage points.

TABLE 1 (CONTINUED)

NECAP scores regressed on different student ACCESS scores and student and school characteristics, 2006

	Panel 2: scale score points					
	Reading		Writing		Math	
Predictor	5th grade	8th grade	5th grade	8th grade	5th grade	8th grade
Intercept[a]	538.8***	833.4**	533.9	830.5**	539.5**	830.7
ACCESS predictors[b]						
Listening	0.1	0.0	0.2	0.3***	0.3***	0.0
Speaking	0.2****	0.1***	0.0	0.2****	0.0	0.0
Reading	1.3****	1.1****	1.3****	0.8****	1.1****	1.0****
Writing	1.0****	1.0****	1.5****	1.2****	0.6****	0.9****
Student characteristics						
Gender[c]	0.2	0.8	2.6****	2.6****	–1.6***	–1.4**
Poverty status[d]	–1.8****	–0.4	–1.1	0.0	–1.3	0.2
Disability status[e]	–4.0****	–5.8****	–4.6****	–4.5****	–4.5****	–6.1****
Age status[f]	–0.9	–1.2	–2.1	–0.6	–0.2	–0.3
Asian	–0.4	0.9	0.1	–1.0	–0.3	2.8**
Non-Hispanic Black	–3.6***	–0.2	–2.4	–1.4	–4.7****	–1.4
Hispanic	–1.6	–1.6	–1.5	–3.6***	–3.2****	–2.3**
Years in English language learner programs	–0.7****	–0.1	–0.7***	0.1	–0.8****	–0.1
School characteristics						
School size[g]	0.2	0.0	0.1	–0.1	0.0	–0.1
School poverty[h]	–0.1	–0.6	0.2	–0.2	–0.6**	–1.4***
Racial composition[i]	–0.1	–0.5	0.3	–0.6	–0.9***	–1.1***
English language learner student density[j]	0.2	–0.4	–0.2	–0.5	0.0	–0.2
Rural	2.7**	1.9	1.7	0.0	1.3	2.4
Urban	–0.3	–1.0	–0.6	–2.2	–0.5	0.4
New Hampshire	–0.8	–1.4	–2.4	–0.7	2.1	0.7
Vermont	0.1	1.4	–2.6	2.8	2.7	3.0

j. Defined as the percent of students in the school who are English language learners. Measured in units of 10 percentage points.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007). For more details, see appendix G.

TABLE 2
Percent of additional and total variance in 5th and 8th grade NECAP scores explained by the three models, 2006

Content area	Grade	Model 1: additional variance explained by student covariates	Model 2: additional variance explained by school covariates	Model 3: variance explained by ACCESS scores	Model 3: total variance explained
Reading	5	17	2	30	49
	8	19	9	23	51
Writing	5	14	2	28	44
	8	18	7	25	50
Mathematics	5	14	2	21	37
	8	17	5	14	36

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007). For more details, see appendix G.

Predictors of NECAP outcomes in reading, writing, and mathematics

This section describes the predicted changes for each NECAP outcome (reading, writing, and mathematics) within each grade, focusing on the absolute and relative predictive strength of each ACCESS subscale domain score. The student and school covariates are also examined as well as the variance in NECAP scores explained in each of the final models.

Reading. NECAP reading scores in both 5th and 8th grades were significantly and positively predicted by ACCESS reading, writing, and speaking scores after controlling for the covariates in model 3. Among the ACCESS domain scores the strongest predictor of NECAP reading outcomes was ACCESS reading scores, followed by ACCESS writing and speaking scores. ACCESS domain scores explained 30 percent of the variance in NECAP reading scores in 5th grade and 23 percent in 8th grade.

NECAP reading scores in both 5th and 8th grades were significantly and positively predicted by ACCESS reading, writing, and speaking scores after controlling for the covariates in model 3

ACCESS reading scores were the strongest predictor among all the ACCESS domain scores of NECAP reading outcomes in 5th and 8th grades, followed by ACCESS writing, speaking, and listening scores after accounting for other variables. In 5th grade the regression coefficient associated with ACCESS reading scores was significantly larger than the regression coefficients for writing, speaking, and listening scores, and the regression coefficient for writing was significantly larger than the regression coefficients for listening and speaking. In 8th grade the regression coefficient for ACCESS reading scores was not significantly different from the regression coefficient for ACCESS writing scores, but both regression coefficients were significantly larger than the ones associated with ACCESS speaking and listening scores. Panel 1 of table 1 shows that a 5th grade English language learner student whose ACCESS reading score was 1 standard deviation higher was predicted to have an NECAP reading score 0.383 standard deviation higher, holding other ACCESS scores and covariates constant. ACCESS reading scores were a significantly stronger predictor of NECAP reading outcomes in 5th grade (0.383 standard deviation) than in 8th grade (0.283 standard deviation).[5] In contrast, the differences between the 5th and 8th grade regression coefficients for ACCESS writing, speaking, and listening scores were not statistically significant.

Among the covariates included in the model, an English language learner student's disability status was a significant predictor of NECAP reading

scores in both grades after controlling for student ACCESS scores and other covariates. Reading outcomes were significantly lower (0.337 standard deviation or 4.0 scale score points in 5th grade and 0.487 standard deviation or 5.8 scale score points in 8th grade) for students with disabilities than for students without disabilities, holding all other variables constant. In addition, 5th grade English language learner students who were non-Hispanic Black were predicted to have significantly lower (0.298 standard deviation or 3.6 scale score points) NECAP reading scores than White students. Holding other variables constant, 5th grade English language learner students who were living in poverty or who had spent an extra year in English language learner programs were predicted to have significantly lower NECAP reading scores by 0.146 standard deviation (1.8 scale score points) or 0.061 standard deviation (0.7 scale score point), respectively.

Table 2 shows that student and school covariates together explain 19 percent of the total variance in 5th grade and 28 percent of the total variance in 8th grade NECAP reading scores. As a group, the ACCESS domain scores explain an additional 30 percent of the total variance in NECAP reading scores in 5th grade and an additional 23 percent of the total variance in 8th grade after controlling for student and school covariates.

Writing. NECAP writing scores were significantly and positively predicted by ACCESS reading and writing scores in 5th grade and by all four ACCESS domain scores in 8th grade after controlling for the covariates in model 3. ACCESS reading and writing scores were the strongest predictors of NECAP writing outcomes for 5th and 8th grades. ACCESS domain scores explained 28 percent of the total variance in NECAP writing scores in 5th grade and 25 percent in 8th grade.

Among 5th grade English language learner students only ACCESS reading and writing scores were significant predictors of NECAP writing outcomes after controlling for other ACCESS scores and covariates. Similar to the results for NECAP reading scores, changes in ACCESS reading scores predicted the largest changes in 5th grade NECAP writing scores compared with changes in other ACCESS scores after controlling for the covariates in the model. The regression coefficient associated with ACCESS reading was significantly larger than the coefficients associated with speaking and listening, but there was no significant difference between the coefficients associated with ACCESS reading and writing scores for predicting 5th grade NECAP writing scores. For every 1 standard deviation increase in ACCESS reading scores, 5th grade ACCESS writing scores were predicted to increase significantly by 0.303 standard deviation.

NECAP writing scores were significantly and positively predicted by ACCESS reading and writing scores in 5th grade and by all four ACCESS domain scores in 8th grade after controlling for the covariates in model 3

Whereas only ACCESS reading and writing scores were significant predictors of NECAP writing outcomes in 5th grade, each of the four ACCESS domain scores was a significant predictor of NECAP writing outcomes in 8th grade after holding other ACCESS domain scores and covariates constant. ACCESS writing scores were the strongest predictor of 8th grade NECAP writing scores: panel 1 of table 2 shows that a 1 standard deviation increase in ACCESS writing scores predicted a 0.268 standard deviation increase in NECAP writing scores. The regression coefficient associated with ACCESS writing was significantly larger than the coefficients associated with both speaking and listening but not significantly larger than the coefficient associated with ACCESS reading.

The standardized regression coefficients show that after controlling for student and school covariates, higher ACCESS reading scores predicted significantly larger increases in NECAP writing scores in 5th grade than in 8th grade. The relationship between ACCESS and NECAP writing scores was also larger in 5th grade (0.294 standard deviation) than in 8th grade (0.268 standard deviation), though the difference was not statistically

significant. The opposite pattern was true for ACCESS listening and speaking scores. ACCESS speaking and listening scores were not significantly related to NECAP writing outcomes in the 5th grade, but they were significantly related in 8th grade. While there was no significant difference between the regression coefficients for ACCESS listening scores in 5th (0.048 standard deviation) and 8th grades (0.110 standard deviation), ACCESS speaking scores were significantly stronger predictors of NECAP writing outcomes in 8th grade (0.135 standard deviation) than in 5th grade (0.015 standard deviation).

After holding ACCESS scores and other covariates constant, predicted NECAP writing scores were significantly higher for girls than for boys in 5th and 8th grades (0.175 standard deviation in 5th grade and 0.199 in 8th grade, or 2.6 scale score points in both grades), and scores were significantly lower (by 0.312 standard deviation in 5th grade and 0.337 in 8th grade, or about 4.5 scale score points in both grades) for students with disabilities than for students without disabilities. Each additional year spent in an English language learner program above the average for all English language learner students in the 5th grade sample was associated with a significantly lower NECAP writing score by 0.047 standard deviation, or 0.7 scale score point. The predicted NECAP writing score in 8th grade was significantly lower (by 0.272 standard deviation or 3.6 points) for Hispanic students than for non-Hispanic White students. Holding all else constant, none of the school covariates were significantly related to NECAP writing outcomes in either grade.

NECAP mathematics scores were positively and significantly predicted by ACCESS reading and writing scores in both 5th and 8th grades after controlling for the covariates in model 3

The four ACCESS scores combined explained an additional 28 percent of the variance in NECAP writing scores in the 5th grade and 25 percent of the variance in 8th grade after controlling for student and school covariates (see table 2). By comparison, the student and school covariates together explained 16 percent of the variance in 5th grade NECAP writing scores and 25 percent in 8th grade scores.

Mathematics. Like NECAP reading and writing scores, NECAP mathematics scores were positively and significantly predicted by ACCESS reading and writing scores in both 5th and 8th grades after controlling for the covariates in model 3. Among the ACCESS domain scores included in the model ACCESS reading scores were the strongest predictor of NECAP mathematics outcomes for both 5th and 8th grade English language learner students, followed by ACCESS writing scores. ACCESS domain scores explained 21 percent of the variance in NECAP mathematics scores in 5th grade and 14 percent in 8th grade.

Table 1 shows that in both grades, after controlling for covariates in the model, NECAP mathematics scores were most strongly predicted by ACCESS reading scores, followed by writing scores and scores in the two oral proficiency domains. In 5th grade the regression coefficient associated with ACCESS reading was significantly larger than the coefficients associated with writing, speaking, and listening. In 8th grade the pattern was similar except that there was no significant difference between the regression coefficients associated with ACCESS reading and writing scores. There was no significant difference between the coefficients for ACCESS reading scores in 5th and 8th grades (0.320 and 0.260 standard deviation, respectively), nor for the coefficients of any of the other ACCESS scores.

In both grades school poverty and racial composition were significant predictors of NECAP mathematics scores after holding ACCESS scores and other covariates constant. Specifically, for every 10 point increase in the percentage of students in the school who were living in poverty or were White, 5th grade NECAP mathematics scores were predicted to be lower by 0.050 standard deviation or 0.078 standard deviation, respectively, after controlling for other variables in the model. A 10 percentage point increase in school poverty levels and racial composition was significantly associated with a decrease in 8th grade NECAP

mathematics scores by 0.115 and 0.087 standard deviation, respectively.

After holding ACCESS scores and student and school covariates constant, 5th grade students with disabilities had a predicted NECAP mathematics score that was 0.384 standard deviation or 4.5 scale score points lower than the score for students without disabilities. Similarly, 8th grade students with disabilities had a predicted NECAP mathematics score that was 0.497 standard deviation or 6.1 scale score points lower than the score for their counterparts without disabilities. In both grades the differences were statistically significant. Hispanic students had predicted NECAP mathematics scores that were significantly lower by 0.275 standard deviation (or about 3 scale score points) in 5th grade and 0.182 standard deviation (or about 2 points) in 8th grade than the scores for non-Hispanic White students. In addition, girls were predicted to have NECAP mathematics scores that were 1.6 scale score points (0.132 standard deviation) lower than the scores for boys in 5th grade and 1.4 scale score points (0.110 standard deviation) lower than the scores for boys in 8th grade. Non-Hispanic Black students had a predicted score that was nearly 5 scale score points (0.395 standard deviation) lower than non-Hispanic White students in 5th grade and 1.4 scale score points (0.115 standard deviation) in 8th grade. Asian students had a predicted NECAP mathematics score almost 3 scale score points (or 0.228 standard deviation) higher than non-Hispanic White students in 8th grade. For both grades more student and school covariates were significant predictors of NECAP mathematics scores than of NECAP reading or writing scores.

In 5th grade student and school covariates together explained 16 percent of the variance in NECAP mathematics scores, and the ACCESS scores explained an additional 21 percent after controlling for student and school covariates. Similarly, in 8th grade the student and school covariates together explained 22 percent of the variance in NECAP mathematics scores, and the ACCESS scores explained an additional 14 percent (see table 2).

Predicted changes across NECAP outcomes for each ACCESS domain

This section compares the predicted changes across NECAP outcomes (reading, writing, and mathematics) and grades within each ACCESS domain, focusing on the predicted scale score point changes only. As noted, although 10 point score shifts are not equivalent across the four ACCESS domains because the scores from each domain have different standard deviations, it is possible to compare the predicted changes in NECAP outcomes associated with 10 scale score point shifts in a single ACCESS domain across NECAP outcomes within a grade.

Reading. ACCESS reading scores were significant predictors of NECAP reading, writing, and mathematics scores in both 5th and 8th grades.

ACCESS reading and writing scores were significant predictors of NECAP reading, writing, and mathematics scores in both 5th and 8th grades

In both 5th and 8th grades a 10 point change in ACCESS reading scale scores predicted similar magnitudes of changes in NECAP reading, writing, and mathematics scale scores. For every 10 point change in 5th grade ACCESS reading scores (holding all other ACCESS scores and covariates constant), there was a significant and positive predicted change of 1.3 points in both NECAP reading and writing scale scores and a 1.1 point change in NECAP mathematics scale scores. In 8th grade for every 10 point change in ACCESS reading scale scores (holding all other ACCESS scores and covariates constant), NECAP scale scores changed 1.1 points for reading, 0.8 points for writing, and 1.0 points for mathematics. Notably, ACCESS reading scores were positive and significant predictors of NECAP mathematics scores in both grades.

Writing. ACCESS writing scores were significant predictors of NECAP reading, writing, and mathematics scores in both 5th and 8th grades.

In contrast to the results for ACCESS reading scores, there was more variation in the predicted changes in NECAP outcomes associated with the ACCESS writing scores in 5th grade. A 10 point change in 5th grade ACCESS writing scale scores predicted a change of 1.0 points in NECAP reading scores, 1.5 points in NECAP writing scores, and 0.6 point in NECAP mathematics scores. In 8th grade for every 10 point change in ACCESS writing scale scores (holding all other ACCESS scores and covariates constant) NECAP scale scores changed 1.0 points for reading, 1.2 points for writing, and 0.9 points for mathematics. As with ACCESS reading scores, ACCESS writing scores were positive and significant predictors of NECAP mathematics scores in both grades.

Speaking and listening. ACCESS speaking and listening scores were significant predictors of NECAP scores for only four outcomes: 5th and 8th grade reading (speaking), 8th grade writing (speaking and listening), and 5th grade mathematics (listening).

A 10 point change in ACCESS listening or speaking scores (holding all other ACCESS scores and covariates constant) had smaller predicted changes in all three NECAP scores than did ACCESS reading and writing scores. The ACCESS speaking score was a significant predictor of 5th and 8th grade reading (0.2 and 0.1 points, respectively) and of 8th grade writing (0.2 points) but was not a significant predictor of NECAP mathematics scores. The ACCESS listening score was a significant predictor of only 8th grade writing (0.3 points) and 5th grade mathematics (0.3 points).

This report finds evidence that ACCESS measures of English literacy skills (reading and writing) have stronger associations with NECAP content outcomes than do ACCESS measures of English oral skills (listening and speaking)

DISCUSSION, FUTURE RESEARCH, AND STUDY LIMITATIONS

After controlling for other ACCESS scores as well as student and school characteristics, ACCESS scores in reading and writing were the strongest predictors of English language learner student performance on the NECAP reading, writing, and mathematics assessments. This report thus finds evidence to support the hypothesis that ACCESS measures of English literacy skills (reading and writing) have stronger associations with NECAP content outcomes than do ACCESS measures of English oral skills (listening and speaking). Of the two English language literacy skills, higher ACCESS reading scores were associated with the largest increases in NECAP outcomes in all three content areas.

English literacy skills (as measured by ACCESS) were positive and significant predictors of NECAP mathematics scores. In school districts subject to federal NCLB regulations, new English language learner students are not required to take large-scale assessments in reading and writing during their first year, but they are required to take large-scale assessments in mathematics. The findings from this report suggest, however, that the English language skills of English language learner students, specifically reading and writing, are strong predictors of NECAP mathematics outcomes. This is similar to the finding for NECAP reading and writing outcomes. While ACCESS scores explain less of the variance in NECAP mathematics scores than in NECAP reading and writing scores, the ACCESS reading score is a stronger predictor of 5th and 8th grade NECAP mathematics scores than of NECAP writing scores in both grades. Further examination of how English language skills are related to mathematics performance is an area for future research.

Additional observations and topics for future research

Looking at patterns between grades, the strength of the relationship between ACCESS reading scores and NECAP reading and writing outcomes in 5th and 8th grades differ significantly, with the relationship in 5th grade stronger. There were no other significant differences across grades except for a significant increase in the size of the coefficient for ACCESS speaking scores in predicting NECAP writing scores. These patterns raise

important questions about how and why English language proficiency in different domains may have varying relationships with content knowledge for English language learner students at different grades.

Consistent with documented national and international trends, girls had significantly higher scores on the NECAP writing assessment and significantly lower scores on the NECAP mathematics assessment than boys did (holding all ACCESS scores and other background characteristics constant). Also consistent with findings from prior research, students with disabilities in the report's English language learner samples received significantly lower scores than did students without disabilities on all three NECAP assessments, again holding other variables constant. Unlike findings from other research, however, school poverty was not a significant predictor of NECAP outcomes except in 5th and 8th grade mathematics.

Two other intriguing findings also emerged. First, among 5th grade English language learner students each additional year spent in English language learner programs was associated with significantly lower NECAP outcomes in all three content areas. Second, among both 5th and 8th grade English language learner students higher proportions of White students in the school were associated with significantly lower NECAP mathematics outcomes. How length of participation in English language learner programs affects English language acquisition and performance on content assessments, as well as how English language learner program types may be related to these outcomes, are rich areas for future study. Similarly, how a school's racial composition affects English language learner performance outcomes may be a complex yet interesting area for additional research.

Finally, because this report is one of the first efforts to investigate and compare the results of two new English language proficiency and large-scale content assessments, many additional research questions can be explored with these data. For example, this report examined ACCESS scores in the four language domains only. However, the ACCESS assessment measures academic English language proficiency in several other areas, such as in the academic language needed for mathematics and science. Future research could examine the relationship between other ACCESS proficiency scores and large-scale assessment outcomes.[6]

Unlike findings from other research, school poverty was not a significant predictor of NECAP outcomes except in 5th and 8th grade mathematics

Because ACCESS and NECAP data are collected for students in New Hampshire, Rhode Island, and Vermont each year, the two datasets also provide opportunities to examine English language acquisition rates and performance on content assessments over time. And detailed examinations of the assessments themselves may provide useful information for both researchers and educators. For instance, are there items on the NECAP assessment that perform differently for new English language learner students versus advanced English language learner students? In addition, examinations of differential item functioning could help elucidate the relationship between language acquisition and performance on content assessments.

Study limitations

This report finds statistically significant correlations between English language learner student scores on two assessments, after controlling for several student and school characteristics. Correlation does not equal causation, however. Causal conclusions cannot be drawn from the findings in this report. English language learner students examined in this report took the ACCESS assessment before taking the NECAP assessments, but this report does not provide evidence that higher English language proficiency in some domains (measured by ACCESS) causes or leads to higher NECAP outcomes. Unmeasured factors (such as student motivation or access to high-quality teachers) may have raised both ACCESS scores and

English language learner students examined in this report took the ACCESS assessment before taking the NECAP assessments, but this report does not provide evidence that higher English language proficiency in some domains causes or leads to higher NECAP outcomes

NECAP outcomes. Causal claims about the impacts of English language proficiency on large-scale assessment outcomes can best be drawn from randomized studies that control for all possible background factors.

There are also limitations to the generalizability of the report's findings. As noted, samples were dominated by English language learner students from Rhode Island. Aggregate results for the combined three-state samples may not accurately reflect relationships between ACCESS and NECAP scores for English language learner students in New Hampshire and Vermont. Model results were derived after statistically controlling for state location, but model findings may not represent relationships between predictors and outcomes for a specific state. Because Rhode Island English language learner students constituted almost two thirds of the 5th and 8th grade samples, sample characteristics and aggregate project findings involving data from all three states combined reflect more of the characteristics and assessment outcomes of English language learner students from Rhode Island. English language learner students from the three REL Northeast and Islands states also are not representative of English language learner students in other parts of the country; the findings from this project therefore cannot be generalized to all states.

In addition, the final 5th grade English language learner student sample contained 1,345 cases from an original 1,582 students, and the final 8th grade sample contained 921 cases from an original 1,090 students. The 15 percent of cases dropped from the 5th grade sample and the 19 percent of cases dropped from the 8th grade sample were students who were missing one or more NECAP scores, ACCESS scores, or pieces of information about their individual or school characteristics. Because new English language learner students were not required to take the NECAP reading or writing assessments, it is possible that students who were missing assessment scores or other data were the newest students and may have weaker English language skills. If these students were dropped from the project samples, the findings of this report are applicable primarily to English language learner students with stronger English language skills or to students who have been in the United States for more than one year.

APPENDIX A
REVIEW OF THE LITERATURE

This appendix reviews current understandings of the relationship between English language proficiency and the demonstration of content knowledge among English language learner students. The challenge of measuring content knowledge among English language learner students is discussed as well as different types of English language skills and their potential impacts on student academic performance. A brief review of previous and current generations of English language proficiency assessments is provided, followed by a discussion of the role that student and school characteristics may play in English language acquisition and performance on content assessments.

Measuring content knowledge among English language learner students

The NCLB Act requires that all students be assessed in "a valid and reliable manner" in English language arts and mathematics (Rabinowitz, Ananda, and Bell 2005, p. 2). But existing content assessments may not provide valid measures of content knowledge among English language learner students. By themselves, content assessments are not designed to identify how or if English language limitations may interfere with an ability to communicate content learning. Language issues are an important concern in measuring academic achievement, because to perform well on large-scale content assessments, English language learner students need to master not only the content assessed, but also the academic English language skills needed to engage in content learning within the classroom and to demonstrate knowledge on formal content assessments (Cummins 1981a).

Although the NCLB Act does not require English language learner students to participate in statewide English language arts assessments during their first year in the United States, there is no similar exemption for statewide mathematics assessments, which are usually administered in English. Even after the first year, current content assessments may not always provide valid results for English language learner students (Abedi, Leon, and Mirocha 2003; Zehler and others 1994). Indeed, research has shown that English language learner students can be penalized by language-dependent mathematics assessments (Brown 2005). Evidence about the relationship between English language skills and performance on content assessments may help educators better assess how much low performance among English language learner students is due to language limitations as opposed to—or in addition to—true difficulties with the academic content.

Types of English language skills needed for academic success

Researchers distinguish social English from the academic English needed to learn academic content (Abella, Urrutia, and Shneyderman 2005). It has long been argued that academic language proficiency is necessary for academic achievement, even as definitions of academic language have varied (Collier 1987; Cummins, 1981a, 1981b; Francis and others 2006; Valdés 2004).[7] Scarcella (2003) argues that social and academic language both draw on knowledge of vocabulary and grammar as well as skills in discourse and higher-order thinking, but academic language requires knowledge of more specialized subject matter vocabulary and specific modes of communication within different media. Similarly, Bailey and Butler (2007) argue that academic language differs from social language by using different vocabularies, types of syntax, and levels of classroom discourse. Academic language involves abstract forms of language needed to communicate in formal, often decontextualized, situations, and may be needed for successful navigation of classroom learning and large-scale assessments.

Researchers and policymakers have also identified four distinct language domains: listening, speaking, reading, and writing. Under the NCLB Act all states must test English language learner

students annually on their English proficiency in these four domains. Previous research has found a link between oral proficiency (listening and speaking combined) and literacy (reading and writing combined) (August and Shanahan 2006). In fact, oral language proficiency is considered to be an essential step in the language acquisition process (Gottlieb 2004). In a review of research on the relationship between oral language proficiency and literacy Geva (2006, p. 139) found that "English oral language proficiency is consistently implicated when larger chunks of text are involved, whether in reading comprehension or writing." In another study Saunders, Foorman, and Carlson (2006) found that oral proficiency had a significant relationship with literacy development.

Although proficiency in English oral skills may be an important foundation for the development of English reading and writing skills, English oral proficiency may not be sufficient for English language learner students to perform well on large-scale content assessments. Gottlieb (2006) has argued that too often teachers equate oral proficiency in social English with readiness for academic English. In addition, Geva (2006, p. 135) has argued that proficiency in reading is dependent on "precursor literacy skills" that do not involve oral proficiency. Because large-scale assessments typically involve academic forms of reading and writing, proficiency in these two language domains may be more important than oral proficiency for strong performance on such assessments.

Little is known about the relationship between language learning and content learning. A longitudinal study on transitional bilingual education, in which students were taught content in their native language while learning English, found that students who were held in a transitional program for a longer period of time (after 5th grade) had higher achievement than those in English-only programs (Ramírez 1992). Students in two-way bilingual programs or dual-language programs learn content in two languages (Lindholm-Leary 2001). Several studies have shown that the academic achievement of these students is comparable in both languages (Cazabon, Nicoladis, and Lambert 1998; DeJong 2002). However, there are still few studies that look in depth at the relationship between language learning and content learning, in part because assessments of English proficiency in the past have not provided sufficient information about students' skills in specific language domains.

Previous and new generations of English language proficiency assessments

Until recently, most measures of English proficiency focused on general language acquisition. Researchers have found that these assessments were "not appropriate for assessing readiness for taking standardized assessments in English" (Stevens, Butler, and Castellón-Wellington 2001, p. 38). Most of these English language assessments did not effectively differentiate between levels of academic English readiness in the different domains of language across and within content areas (Abedi and Lord 2001). Furthermore, traditional English language assessments were not helpful in reclassifying students out of the English language learner category, and concerns existed about the reliability and validity of the assessments (Abedi, Leon, and Mirocha 2003; Mahoney, Haladyna, and MacSwan 2006; Zehler and others 1994). Studies comparing scores on English language assessments and large-scale content assessments therefore failed to provide educators with information to help English language learner students become proficient in the academic language required to succeed in mainstream classrooms and on statewide assessments. One study comparing results from English language proficiency assessments and statewide content assessments found correlations between the two assessments, but because the study used older English language proficiency assessments, it was unable to identify the academic language constructs being measured (Albus and others 2004).

A new generation of English language proficiency assessments has been designed explicitly to measure the academic language required for success

on content assessments. English language learner assessment experts agree that newly developed English language proficiency assessments may be better aligned with standards-based instruction and large-scale content assessments (Butler and others 2004; Gottlieb 2003; Mahoney and MacSwan 2005). The ACCESS assessment is one of the new English language proficiency assessments that have been designed to measure proficiency in different English language domains (see appendix D for details about the assessment). Data from this assessment have the potential to provide educators with rich information on multiple dimensions of English language proficiency and how proficiencies in different domains may be related to achievement on content assessments.

The role of student and school characteristics in assessment outcomes

Studies that seek to determine whether and how performance on English language proficiency assessments may be related to outcomes on content assessments should take into account other student and school factors that may be correlated with both sets of scores. Although there is little research that focuses specifically on the impact of student characteristics on the acquisition of English, there is ample research that links individual characteristics, such as gender, disability status, race/ethnicity, and poverty status, to large-scale assessment outcomes (Lee, Grigg, and Dion 2007; Lee, Grigg, and Donahue 2007). In the 2007 National Assessment of Educational Progress 8th grade reading assessment, students with disabilities scored significantly lower than students without disabilities, as did students eligible for free or reduced-price lunch (Lee, Grigg, and Donahue 2007). Scores also varied by race/ethnicity, with non-Hispanic Black students and Hispanic students scoring lower than White and Asian students. Gender differences are a bit more complex, with boys, on average, performing better in mathematics, and girls, on average, performing better in reading and writing (Cole 1997; Coley 2001; Freeman 2004; Klecker 2006; Meadows, Land, and Lamb 2005; Nowell and Hedges 1998).

Other factors may also be related to academic achievement and performance on content assessments among English language learner students. Retention of English language learner students is of particular concern, given research demonstrating that it takes 6–10 years to become proficient in English reading and writing (Thomas and Collier 2002). The number of English language learner students being retained in the last six years has increased (Solórzano 2008), with no evidence that these older students demonstrate improved learning. Thus, research on English language learner performance should include information about whether students are old for their grade. Similarly, in a study of Spanish- and Vietnamese-speaking elementary school students, Hakuta (2000) found that it took five years for 90 percent of the students to be proficient in oral English, and seven years for 90 percent of them to be proficient in reading and writing. Thus, the number of years a student has been participating in English language learner programs may be associated with their performance on content assessments.

School characteristics—in particular, school poverty and school size—have also been found to be associated with large-scale assessment outcomes (Lee 2000; Ma and Wilkins 2002). Minority students in general may experience a negative impact when in a classroom with a high percentage of students eligible for free or reduced-price lunch, regardless of their own status (Muijs and Reynolds 2003). Students attending more racially segregated schools tend to have lower assessment scores (Bifulco and Ladd 2007). Some studies have found no significant relationship between school size and achievement at the elementary and high school levels (Gardner 2001), while others have found that smaller elementary and high schools tend to have higher achievement (Caldas 1993; Fowler and Walberg 1991; McMillen 2004). Thus, school characteristics such as school poverty, school racial composition, and school size should be considered when examining the relationship between English language proficiency and performance on content assessments.

APPENDIX B
METHODS OF ANALYSIS

Multilevel regression models were fit to the 5th and 8th grade English language learner student data to predict NECAP outcome variables using ACCESS scores and student and school covariates. Because regression analysis and observational data were used, the estimated relationships represent partial correlations and do not imply causation. Rather, the regression coefficients in the models describe the association between a dependent variable (for example, one of the NECAP scores) and the independent variables (ACCESS scores) holding all other covariates (student and school characteristics) in the model constant.

Multilevel regression models were used to account for the interdependence of assessment scores among English language learner students attending the same schools. The percent of total variation in NECAP outcome scores between schools (the intraclass correlation coefficient) was significant in all content areas and for both grades. In 5th grade the intraclass coefficient was 15.5 percent in reading, 20.5 percent in writing, and 13.0 percent in mathematics. In 8th grade it was 26.8 percent in reading, 28.2 percent in writing, and 16.4 percent in mathematics.

The multilevel regression models were fit to the 5th and 8th grade English language learner student data in stages. First, NECAP reading, writing, and mathematics scores were regressed on the student and school covariates (models 1 and 2). Then ACCESS scores were added (model 3). Thus, in addition to the unconditional model that included only a random school effect, three models were fit to the 5th and 8th grade data samples for each NECAP outcome variable (reading, writing, and mathematics scores).

Estimates of relationships between outcomes and predictors

The regression coefficients in each model represent the predicted change in NECAP scores for every unit change in a predictor or covariate while holding all other variables in the model constant. The coefficients therefore provide an estimate of the strength of the relationship between a specific NECAP outcome and an ACCESS domain score or covariate while controlling for other variables. The regression coefficient estimates were reported two ways: in standard deviation units and in scale score points.

In standard deviation units. NECAP and ACCESS scores were each standardized to have a mean of 0.0 and a standard deviation of 1.0 before models were fit to the data. Therefore, the regression coefficients associated with the ACCESS scores were estimated in standard deviation units. The regression coefficients for the ACCESS predictors represent standard deviation changes in NECAP scores for every one standard deviation change in an ACCESS domain score, holding all other variables constant. Regression coefficients for the student and school covariates represent standard deviation changes in NECAP scores for every unit change in the covariate (defined in table C6 in appendix C).

In scale score points. To aid in interpreting the predicted NECAP score changes measured in standard deviation units, the regression coefficients were converted back to their point values on the original scale and are presented alongside the standardized estimates. Standardized ACCESS scores were also converted back to their point values on the original scale, and predicted changes in NECAP scale scores were calculated for 10 point changes in point values for each ACCESS domain.[8] Readers are cautioned not to compare the predicted changes in NECAP outcomes associated with 10 scale score point shifts in ACCESS domain scores within a single model. Ten point score shifts are not equivalent across the four ACCESS domains because the scores from each domain have different standard deviations.[9]

However, it is possible to compare the predicted changes in NECAP outcomes associated with 10 scale score point shifts in a single ACCESS domain across NECAP outcomes within a grade. For example, it is possible to compare the predicted

changes in NECAP reading, writing, and mathematics score associated with a 10 scale score point shift in ACCESS reading scores at the 5th grade level. Predicted changes in NECAP scale scores were calculated for unit changes in the covariates as defined in table C6 in appendix C.

Two sets of statistical tests were also conducted and reported (see appendix E). When the .95 confidence intervals constructed around individual regression coefficients did not include zero, coefficients were reported as statistically significant (different from zero). When .95 confidence intervals constructed around the difference between two standardized regression coefficients did not include zero, the larger coefficient was reported as significantly "stronger" than the other coefficient. In this report "stronger" predictors are defined as those whose regression coefficients are simply larger than those of other noted predictors in the report's regression models.[10]

Estimates of variance explained

The multilevel regression models also generated estimates of the total percentage of variance in the NECAP outcome measures that was explained by the student and school covariates and the ACCESS domain scores. This percentage is analogous to R^2 in a traditional ordinary least squares model in which higher percentages of explained variance are associated with stronger prediction models.

Appendix F provides additional details on the multilevel modeling procedures used and the calculation of the percentage of variance explained. Appendix G presents the results of the multilevel models in which NECAP scores in reading, writing, and mathematics are regressed on the student and school covariates and the ACCESS domain scores for both the 5th and 8th grade samples.

APPENDIX C
ABOUT THE DATA

Data from several sources were merged to create the datasets examined in this report, and data were imputed for numerous cases with missing assessment scores. The data sources and procedures used are presented below.

Data sources and merging procedures

The datasets examined in this report were created by merging records from three sources: student English language learner scores and demographic data from the ACCESS assessment, student English language learner scores and demographic data from the NECAP assessment, and school data from the Common Core of Data of the National Center for Education Statistics. The data merged were for English language learner students and the schools they attended in New Hampshire, Rhode Island, and Vermont. ACCESS data were received from the WIDA consortium, the developers of the ACCESS assessment, and NECAP data were received from Measured Progress, the developers of the NECAP assessment. Data on the school characteristics of English language learner students in the three states were downloaded from the Common Core of Data web site.

Using these three data sources, two primary datasets of English language learner student assessment and demographic information were created. The first (called the 5th grade English language learner student sample) contained ACCESS scores and demographic data for English language learner students in 4th grade in spring 2006 and content assessment and demographic data for the same students with 5th grade NECAP data in fall 2006.[11] The second (called the 8th grade English language learner student sample) contained ACCESS data gathered from English language learner students in 7th grade in spring 2006 and assessment and demographic data for the same students with 8th grade NECAP data in fall 2006.

ACCESS and NECAP data records were merged for each state using common student identification numbers within each dataset. For New Hampshire and Vermont school data from the Common Core of Data were merged with the ACCESS and NECAP data by matching school and district names contained within the three databases. For Rhode Island school identification codes used in the Common Core of Data database were retrieved from the Rhode Island Department of Education web site. These school codes were merged into the ACCESS–NECAP datasets by school name, and the codes were then used to merge school Common Core of Data information with Rhode Island English language learner student records in the ACCESS-NECAP datasets.

Not all the 1,582 English language learner students in New Hampshire, Rhode Island, and Vermont who were recorded as 4th graders in the ACCESS database in spring 2006 had 5th grade NECAP data the following fall (table C1). Almost 10 percent did not have any NECAP scores and were therefore dropped from the sample. Of the 1,429 who had records in the NECAP database, 1,388 (97 percent of students with NECAP data, or 88 percent of the original sample) were recorded as having 5th grade NECAP data the following fall. About 3 percent of English language learner students with NECAP data took NECAP assessments designed for other grades; these records were dropped from the 5th grade English language learner student sample.

Similar procedures were followed to create the 8th grade English language learner student sample. Of the 1,090 English language learner students in the three states who were recorded as 7th graders in the 2006 ACCESS database (table C2), 10 percent were missing NECAP data and were dropped from the sample. Of the remaining English language learner students who were recorded as having taken the NECAP, 921 (94 percent of students with NECAP data, or 84 percent of the original sample) were recorded with 8th grade NECAP data. These records were kept for the 8th grade English language learner sample, and all others were dropped.

Imputation methods

Within the 5th and 8th grade samples not all English language learner student records

TABLE C1

NECAP data for English language learner students with 4th grade ACCESS data, 2006

Dataset	Number	Percent
Total	1,582	100.0
Missing NECAP scores	153	9.7
With NECAP data[a]	1,429	90.3
NECAP test level recorded		
Grade 3	3	0.2
Grade 4	26	1.6
Grade 5	1,388	87.7
Grade 6	12	0.8

a. Not all students who took the 4th grade ACCESS assessment in spring 2006 were recorded as having taken the 5th grade NECAP assessment in fall 2006.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005) and student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006).

TABLE C2

NECAP data for English language learner students with 7th grade ACCESS data, 2006

Dataset	Number	Percent
Total	1,090	100.0
Missing NECAP scores	109	10.0
With NECAP data	981	90.0
NECAP test level recorded		
Grade 5	16	1.5
Grade 6	2	0.2
Grade 7	42	3.9
Grade 8	921	84.5

a. Not all students who took the 7th grade ACCESS assessment in spring 2006 were recorded as having taken the 8th grade NECAP assessment in fall 2006.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005) and student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006).

contained complete data. The 5th grade sample contained 1,388 cases, but 67 (5 percent) were missing NECAP, ACCESS, or student or school background data (table C3).[12] Of these cases, 48 were missing one or more NECAP scores, 8 were missing one or more ACCESS scores, and 12 were missing student or school background data.[13] The 8th grade sample contained 921 cases, but 86 (10 percent) were missing one or more assessment scores or student or school background data (table C3).

Missing NECAP and ACCESS scores for 24 cases in the 5th grade sample and 51 cases in the 8th grade sample were imputed using existing NECAP and ACCESS scores and stochastic regression imputation procedures (Little and Rubin 1987). For both datasets NECAP mathematics scores (which were missing least often within the NECAP data) were used to impute missing NECAP reading and writing scores. NECAP reading scores (which were missing less often than NECAP writing scores) were used to impute missing NECAP mathematics scores. Missing ACCESS scores were imputed with other ACCESS scores, using available scores from the language domains that would maximize the number of cases that could be imputed.

For example, missing 5th grade ACCESS listening scores were imputed by regressing available 5th grade ACCESS listening scores on 5th grade ACCESS speaking and writing scores. When both ACCESS speaking and writing scores were available for a specific student record, the imputed ACCESS listening score was the value predicted from the regression model, plus the value of a random error term. The error term was randomly selected from a distribution of possible error terms with a mean of 0 and a standard deviation equal to the standard deviation of the residuals around the estimated regression line.

Because few student and school variables were missing data in both datasets and because these variables were not the primary predictors or outcomes, the research team did not impute data for these background characteristics. After imputing as many missing NECAP and ACCESS scores as the available data allowed, the research team dropped all cases with any remaining missing data. The final 5th grade sample contained 1,345 records, representing 97 percent of all cases with 4th grade ACCESS and 5th grade NECAP data. The final 8th grade sample contained 886 records, representing 96 percent of all cases recorded with the corresponding grade-level assessment data.

TABLE C3

5th and 8th grade English language learner dataset, before and after imputation and deletion of cases with missing data, 2006

Dataset	5th grade			8th grade			Variable used for imputation
	Number of data cases		Number of imputed cases	Number of data cases		Number of imputed cases	
	Before imputation	After imputation		Before imputation	After imputation		
Total, including cases with missing data	1,388	1,388	na	921	921	na	
Total, excluding cases with missing data[a]	1,321	1,345	na	835	886	na	
Cases with missing data[b]	67	43	24	86	35	51	
Cases with missing data for one or more NECAP content area[b]	44	17	27	44	17	27	
Math	26	25	1	13	10	3	NECAP reading
Reading	45	25	20	37	10	27	NECAP mathematics
Writing	47	26	21	44	17	27	NECAP mathematics
All three content areas[c]	25	25	0	10	10	0	
Cases with missing data for one or more ACCESS subject[b]	38	12	26	38	12	26	
Listening	15	9	6	15	9	6	ACCESS speaking (5th and 8th grades) and writing (8th grade)
Speaking	24	7	17	24	7	17	ACCESS listening (5th grade), reading (5th and 8th grades), and writing (8th grade)
Reading	12	8	4	12	8	4	ACCESS speaking (5th and 8th grades) and writing (8th grade)
Writing	15	8	7	15	8	7	ACCESS listening (5th and 8th grades), reading (5th grade), and speaking (8th grade)
All four subjects[c]	6	6	0	6	6	0	
Cases with missing data for one or more background characteristics[b]	9	9	0	9	9	0	
Student age status	2	2	0	2	2	0	
Race	6	6	0	6	6	0	
All school variables	1	1	0	1	1	0	

na is not applicable

a. Listwise deletion was used for student cases missing any ACCESS, NECAP, or background characteristic data.

b. Subcategories may not sum to total because cases were missing data for multiple variables before and after imputing missing data. Cases without complete data were dropped from the sample.

c. Some students were recorded as having taken the NECAP, but scores for all three content areas (reading, writing, and math) were missing.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

Sample characteristics

The students in the 5th and 8th grade English language learner student samples were predominantly Hispanic, living in poverty, and from Rhode Island (table C4). These students were also concentrated in larger, high-poverty, urban schools with large proportions of non-White students and other English language learner students. The demographic characteristics of the English language learner students in the two samples differ from the characteristics of the total student population across the three states in several ways. Hispanic students make up 7.3 percent of students across the three REL Northeast and Islands states but made up 61 percent of the 5th grade English language learner student sample and 59 percent of the 8th grade sample.[14] Whereas 25 percent of all students in the three states were in poverty, the poverty share was 71 percent for the 5th grade English language learner student sample and 63 percent for the 8th grade sample. And whereas 16 percent of all schools in the three states were in urban areas, 67 percent of the 5th grade English language learner student sample and 60 percent of the 8th grade sample were classified as urban.

The New Hampshire, Rhode Island, and Vermont English language learner students in the 5th and 8th grade samples also differed from each other. The students in New Hampshire and Vermont were much more likely than those in Rhode Island to be White or Asian, not living in poverty, and in schools with student populations over 75 percent White and less than 5 percent English language learner students. In both the 5th and 8th grade samples larger shares of English language learner students attended rural schools in Vermont (40 percent) than in New Hampshire (19 percent) and Rhode Island (0.3 percent). The proportions of students who had been in English language learner programs for 5–9 years were also greater in Vermont. Within that state, 60 percent of 5th grade and 73 percent of 8th grade English language learner students had spent at least 5 years in English language learner programs, compared with 32 percent of 5th grade and 39 percent of 8th grade English language learner students in New Hampshire and 9 percent of 5th grade and 38 percent of 8th grade English language learner students in Rhode Island, where English language learner students were more likely to be Hispanic, in poverty, and attending schools in urban areas.

English language learner students from New Hampshire and Vermont differ in demographic characteristics from the national average. The proportion of English language learner students in the student population is 1.5 percent in New Hampshire, 2.5 percent in Vermont, and 6 percent in Rhode Island, whereas English language learner students are 10 percent of the student population across the country (Kohler and Lazarín 2007). Over 67 percent of English language learner students in the 5th grade sample and 60 percent in the 8th grade sample attended schools in urban areas, whereas 91 percent nationwide attend urban schools. Just over 58 percent of 5th grade English language learner students and almost 36 percent of 8th grade students in this report attended schools where more than 20 percent of students were English language learner students. In contrast, 53 percent of English language learner students across the country attended schools with more than 30 percent of other English language learner students (Kohler and Lazarin 2007).

Outcome measures

As noted, the outcome measures examined were 5th and 8th grade scale scores for reading, writing, and mathematics from the 2006 NECAP. The primary predictors were 4th and 7th grade ACCESS scaled scores in English language listening, speaking, reading, and writing. ACCESS scaled scores ranged from 100 to 600. NECAP scaled scores ranged from 500 to 580 for 5th grade and from 800 to 880 for 8th grade.[15] To compare ACCESS scores in the four English language domains and their relationships with NECAP reading, writing, and mathematics scores, all ACCESS and NECAP variables were standardized before they were incorporated into the multilevel models. Table C5 describes the variables, their original scales, and how they were recoded and rescaled prior to inclusion in the multilevel regression models.

TABLE C4

Characteristics of English language learner students from New Hampshire, Rhode Island, and Vermont in the 5th and 8th grade samples, 2006

	5th grade				8th grade			
Demographic	New Hampshire	Rhode Island	Vermont	Total	New Hampshire	Rhode Island	Vermont	Total
Total students (*N*)	350	861	134	1,345	213	578	95	886
Gender								
Male	50.3	50.3	48.5	50.1	57.3	49.3	49.5	51.2
Female	49.7	49.7	51.5	49.9	42.7	50.7	50.5	48.8
Race/ethnicity[a]								
Non-Hispanic Asian	22.6	6.5	29.1	12.9	19.7	9.3	40.0	15.1
Non-Hispanic Black	13.4	6.7	11.2	8.9	15.5	9.0	9.5	10.6
Hispanic	39.4	78.3	9.7	61.3	37.1	74.4	15.8	59.1
Non-Hispanic White	24.6	8.4	48.5	16.6	26.8	7.3	33.7	14.8
Poverty status[b]								
In poverty	56.0	81.1	50.7	71.5	54.5	66.3	62.1	63.0
Not in poverty	44.0	18.9	49.3	28.5	45.5	33.7	37.9	37.0
Disability status[c]								
With disabilities	11.4	17.5	3.0	14.5	9.9	18.7	6.3	15.2
Without disabilities	88.6	82.5	97.0	85.5	90.1	81.3	93.7	84.8
Age status[d]								
Overage	10.6	8.2	9.7	9.0	16.9	13.7	8.4	13.9
Not overage	89.4	91.8	90.3	91.0	83.1	86.3	91.6	86.1
Years in English language learner programs								
0–2	35.4	28.9	27.6	30.5	38.0	42.6	12.6	38.3
3–4	32.6	61.8	11.9	49.2	22.5	19.2	14.7	19.5
5–9	32.0	9.3	60.4	20.3	39.4	38.2	72.7	42.2
Total number of students at school								
Fewer than 300	19.4	25.1	38.1	24.9	2.3	2.6	16.8	4.1
300–499	30.3	51.7	38.8	44.8	8.9	7.6	51.6	12.6
500 or more	50.3	23.2	23.1	30.3	88.7	89.8	31.6	83.3
School location								
Rural	18.9	0.3	40.3	9.1	18.3	0.3	40.0	8.9
Suburban	20.0	23.3	33.6	23.5	25.4	32.9	29.5	30.7
Urban	61.1	76.3	26.1	67.4	56.3	66.8	30.5	60.4
Share of school population receiving free or reduced-price lunch								
Less than 25 percent	39.7	4.8	40.3	17.4	36.6	5.9	45.3	17.5
25–49 percent	33.1	7.0	35.8	16.7	63.4	10.6	46.3	27.1
50 percent or more	27.1	88.3	23.9	65.9	0.0	83.6	8.4	55.4
White student share of school population								
0–24 percent	0.0	73.9	0.0	47.3	0.0	61.8	0.0	40.3
25–49 percent	12.9	8.0	0.0	8.5	0.0	11.9	0.0	7.8
50–74 percent	20.0	8.1	11.2	11.5	0.0	14.0	0.0	9.1
75–100 percent	67.1	10.0	88.8	32.7	100.0	12.3	100.0	42.8

(CONTINUED)

TABLE C4 (CONTINUED)

Characteristics of English language learner students from New Hampshire, Rhode Island, and Vermont in the 5th and 8th grade samples, 2006

	5th grade				8th grade			
Demographic	New Hampshire	Rhode Island	Vermont	Total	New Hampshire	Rhode Island	Vermont	Total
English language learner student density[e]								
Less than 5 percent	34.0	4.9	49.3	16.9	51.2	8.3	62.1	24.4
5–9 percent	15.7	3.8	8.2	7.4	34.3	15.2	2.1	18.4
10–19 percent	28.3	12.2	20.1	17.2	14.6	21.5	35.8	21.3
20 percent or more	22.0	22.4	79.1	58.6	0.0	55.0	0.0	35.9

a. English language learner students of other race/ethnicity are not included because of low numbers.

b. Students in poverty are defined as those who were eligible for free or reduced-price lunch.

c. Students with disabilities are defined as those with Individualized Education Programs.

d. Students who are overage are defined as those exceeding the modal age within the student's grade level by more than 1 year.

e. Share of English language learner students in the school population.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

TABLE C5

Model variables and their scales

Variable	Values	Notes
Outcomes		
NECAP reading, writing, and mathematics	Scale scores were designed to range from 500 to 580 for 5th grade and 800 to 880 for 8th grade.	Scores were converted to standard deviation units. See table C6 for summary statistics of the original variables. In the original scale score metric proficiency in each content area was designated as 540 for 5th graders and 840 for 8th graders.
Primary predictors		
ACCESS listening, speaking, reading, and writing	Scale scores designed to range from 100 to 600.	Scores were standardized to standard deviation units. See table C6 for summary statistics of the original variables.
Student covariates		
Gender	0=Male 1=Female	
Asian	0=No 1=Yes	
Non-Hispanic Black	0=No 1=Yes	
Hispanic	0=No 1=Yes	
Non-Hispanic White	0=No 1=Yes	
Poverty status	0=Not in poverty 1=In poverty	Students in poverty were defined as those who were eligible for free or reduced-price lunch.
Disability status	0=Without disabilities 1=With disabilities	Students with disabilities were defined as those with an Individualized Education Program.
Age status	0= Not overage 1= Overage	Students were defined as overage if they were more than 1 year older than the modal age for their grade level at the time they took the ACCESS assessment.

(CONTINUED)

TABLE C5 (CONTINUED)
Model variables and their scales

Variable	Values	Notes
Years in English language learner programs	0–9 years, centered	The variable was centered around the grand mean in the 5th and 8th grade samples. See table C6 for summary statistics of the original variable.
School covariates		
School size	31–1,557 students, rescaled, centered	The variable was rescaled to units of 100 students and centered around the grand mean in the 5th and 8th grade samples. See table C6 for summary statistics of the original variable.
School poverty	0–100 percent, rescaled, centered	School poverty was defined as the share of students eligible for free or reduced-price lunch. The variable was rescaled to units of 10 percentage points and centered around the grand mean in the 5th and 8th grade samples. See table C6 for summary statistics of the original variable.
Racial composition	0–100 percent, rescaled, centered	School racial composition was defined as the percent of students who were White. The variable was rescaled to units of 10 percentage points and centered around the grand mean in the 5th and 8th grade samples. See table C6 for summary statistics of the original variable.
English language learner student density	0–100 percent, rescaled, centered	English language learner student density was defined as the percent of the school population that was English language learner students. The variable was rescaled to units of 10 percentage points and centered around the grand mean in the 5th and 8th grade samples. See table C6 for summary statistics of the original variable.
Geographic location		
Rural	0=No 1=Yes	
Suburban	0=No 1=Yes	
Urban	0=No 1=Yes	
State		
New Hampshire	0=No 1=Yes	
Rhode Island	0=No 1=Yes	
Vermont	0=No 1=Yes	

Note: English language learner students who were non-Hispanic White, in suburban schools, and attending schools in Rhode Island were the omitted or base-case comparison groups in the models.

Source: Authors' construction.

TABLE C6

Summary statistics of continuous variables used in models, by grade, 2006

	5th grade (*N*=1,345)				8th grade (*N*=886)			
	Mean	Standard deviation	Minimum	Maximum	Mean	Standard deviation	Minimum	Maximum
NECAP scale scores								
Reading	535.5	12.0	500	580	830.6	12.0	800	870
Writing	531.4	14.8	500	580	827.4	13.3	800	877
Math	535.1	11.8	500	570	829.0	12.4	800	862
ACCESS scale scores								
Listening	357.0	42.7	100	484	376.6	50.1	127	471
Speaking	344.5	70.0	121	484	364.3	79.2	139	427
Reading	335.9	35.7	164	436	350.8	31.1	189	445
Writing	330.4	28.8	217	402	345.5	30.3	224	412
Student covariate								
Years in English language learner program	3.3	1.6	0.0	9.0	3.7	2.5	0.0	9.0
School covariates								
School size	451	196	31	1,392	766	278	43	1,557
School poverty[a]	61.3	29.0	0.0	97.7	53.0	25.3	0.2	94.4
Racial composition[b]	45.9	35.1	2.0	100.0	50.3	36.6	0.0	100.0
English language learner student density[c]	25.6	18.4	0.0	84.0	16.5	15.6	0.0	100.0

a. Defined as the percentage of students receiving free for reduced-price lunch.

b. Defined as the percentage of students who are White.

c. Defined as the percentage of students who are English language learners.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

APPENDIX D
DESCRIPTIONS AND RELIABILITY ESTIMATES FOR NEW ENGLAND COMMON ASSESSMENT PROGRAM AND ASSESSING COMPREHENSION AND COMMUNICATION IN ENGLISH STATE-TO-STATE

The Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) English language proficiency assessment examines academic language skills based on English language proficiency standards in different content areas (Gottlieb 2004). The World-Class Instructional Design and Assessment (WIDA) Consortium (a partnership of 14 states and Washington, DC, funded through the federal Enhanced Assessment Grants) developed ACCESS specifically to assess academic English proficiency, that is, the academic language that is required for success in various content areas and in the social and instructional setting of the school. ACCESS for ELLs was piloted in 2005 with 10,000 students across the consortium jurisdictions (10 at the time), and analyses of the assessment yielded high internal reliability (ACCESS for ELLs 2005). The assessment measures English language proficiency in the following ways:

- *By language domain.* ACCESS includes scores for the four language domains (listening, speaking, reading, and writing) and four composite scores that measure literacy, comprehension, oral skills, and composite skills.

- *By content area.* ACCESS measures the academic language skills in each of the four language domains (noted above) in mathematics, English language arts, science, and social studies, as well as the social and instructional setting of the school. Students receive a breakdown of scores in each domain for content areas. For example, they receive a score for the academic speaking skills needed for mathematics and science, allowing a comparison of scores on state mathematics assessments with the ACCESS speaking score for mathematics and science. ACCESS for ELLs does not measure the content itself but student knowledge of the academic language needed for the content areas.

- *By level.* ACCESS covers and measures English language proficiency at five different levels for each grade cluster (K–2, 3–5, 6–8, 9–12), indicating student academic proficiency in each domain from 1 (entering) to 5 (bridging).

Thus, the results of the assessment give teachers and administrators a far more detailed evaluation of each student's English language readiness in various academic areas than previous English language proficiency assessments could. For the first time data are available to address specific questions about English language learner students' language development and academic performance. Rather than simply looking at the relationship between a single English proficiency score and a single large-scale assessment score, ACCESS results allow a deeper, more nuanced analysis of different levels of language proficiency in the content areas in various domains and their relationship to large-scale assessment results.

According to the Annual Technical Report for ACCESS for ELLs (Kenyon and others 2007), the ACCESS assessments provided reliable estimates of students' ability in the domains of listening, reading, speaking, and writing. The assessments in each domain span grade clusters (K–2, 3–5, 6–8, 9–12), so the reliability estimates (Cronbach's α) were calculated using the same clusters. Table D1 presents these reliability estimates. With the exception of the estimate for the listening domain for the grades 3–5 cluster, the reliability estimates are higher than optimal (> 0.70). The reliabilities for the ACCESS assessments were not broken out by demographic subgroups.[16]

New Hampshire, Rhode Island, and Vermont use ACCESS for ELLs as part of their statewide assessment systems. In addition to using ACCESS, the three states also use the same large-scale assessment, the New England Common Assessment Program (NECAP). The NECAP Technical Report

TABLE D1
Reliability estimates for ACCESS subscale scores

Grade cluster	Domain	Weight	Variance	Reliability estimate (α)
Grades 3–5	Listening	0.15	1,455,152	0.68
	Reading	0.15	932,440	0.81
	Speaking	0.35	5,295,808	0.93
	Writing	0.35	850,673	0.89
Grades 6–8	Listening	0.15	1,923,155	0.71
	Reading	0.15	849,777	0.76
	Speaking	0.35	7,598,769	0.94
	Writing	0.35	979,488	0.86

Source: Kenyon and others 2007.

(Measured Progress 2006) describes the collaboration among New Hampshire, Rhode Island, and Vermont that led to the development of assessments for grades 3–8. That report offers three purposes of the assessments:

- To provide data on student achievement in reading and language arts and mathematics to meet the requirements of the NCLB Act.
- To provide information to support program evaluation and improvement.
- To provide parents and the general public information on the performance of students and schools.

According to the NECAP Technical Report, "The tests are constructed to meet rigorous technical criteria, include universal design elements and accommodations so that students can access test content, and gather reliable student demographic information for accurate reporting" (Measured Progress 2006, p. 4).

The three NECAP states emphasize universal design in test development, review all items for bias, and allow all students—including students with disabilities, English language learner students, and general education students—to have access to any of the allowable accommodations. The three states have agreed on 31 accommodations in four areas (alternative settings, scheduling and timing, presentation formats, and response formats). The states chose not to develop a translation of the content assessments, but English language learner students are allowed to use a word-to-word nonelectronic translation dictionary, with no definitions, in mathematics and writing (but not reading) (Measured Progress 2006).

According to the NECAP Technical Report, the NECAP assessments provided reliable estimates of students' ability. Table D2 summarizes the reliability estimate (Cronbach's α) and standard error of measurement for the 5th and 8th grade assessments in mathematics, reading, and writing. The reliability estimates are high (> 0.70).

TABLE D2
Population reliability estimates for NECAP outcome measures

Grade	Content area	Reliability estimate (α)	Standard error of measurement
Grade 5	Mathematics	0.917	4.055
	Reading	0.893	2.988
	Writing	0.750	2.585
Grade 8	Mathematics	0.915	3.893
	Reading	0.897	3.054
	Writing	0.760	2.993

Source: Measured Progress 2006.

TABLE D3
English language learner student subgroup reliability estimates for NECAP outcome measures

Grade	Content area	Reliability estimate (α)
Grade 5	Mathematics	0.90
	Reading	0.89
	Writing	0.77
Grade 8	Mathematics	0.90
	Reading	0.90
	Writing	0.80

Source: Measured Progress 2006.

The reliabilities for the NECAP assessments were also broken out by demographic subgroups. The NECAP developers caution readers that the reliabilities for subgroups depend on the number of individuals in that subgroup; therefore, for smaller subgroups the reliability estimate may be artificially attenuated. Recognizing this limitation, the reliability estimates for the English language learner student subgroup is presented in table D3. These estimates are based on all NECAP assessment takers.

APPENDIX E
CONFIDENCE INTERVALS FOR TESTING DIFFERENCES

TABLE E1

0.95 confidence interval around regression coefficients, by grade level and NECAP content area (within models), 2006

ACCESS subdomain	β	Standard error	ACCESS subdomain	β	Standard error	Difference between βs	Standard error of difference between βs	.95 confidence interval[a] Lower boundary	.95 confidence interval[a] Upper boundary
5th grade NECAP reading									
Reading	0.383	0.03	Writing	0.239	0.03	0.144	0.04	0.061	0.227**
Reading	0.383	0.03	Speaking	0.093	0.03	0.290	0.04	0.213	0.367**
Reading	0.383	0.03	Listening	0.031	0.03	0.352	0.04	0.272	0.432**
Writing	0.239	0.03	Speaking	0.093	0.03	0.146	0.04	0.069	0.223**
Writing	0.239	0.03	Listening	0.031	0.03	0.208	0.04	0.128	0.288**
Listening	0.031	0.03	Speaking	0.093	0.03	−0.062	0.04	−0.136	0.012
5th grade NECAP writing									
Reading	0.303	0.03	Writing	0.294	0.03	0.009	0.04	−0.077	0.095
Reading	0.303	0.03	Speaking	0.015	0.03	0.288	0.04	0.207	0.369**
Reading	0.303	0.03	Listening	0.048	0.03	0.255	0.04	0.172	0.338**
Writing	0.294	0.03	Speaking	0.015	0.03	0.279	0.04	0.198	0.360**
Writing	0.294	0.03	Listening	0.048	0.03	0.246	0.04	0.163	0.329**
Listening	0.048	0.03	Speaking	0.015	0.03	0.033	0.04	−0.045	0.111
5th grade NECAP mathematics									
Reading	0.320	0.03	Writing	0.147	0.03	0.173	0.05	0.082	0.264**
Reading	0.320	0.03	Speaking	0.028	0.03	0.292	0.04	0.207	0.377**
Reading	0.320	0.03	Listening	0.105	0.03	0.215	0.05	0.126	0.304**
Writing	0.147	0.03	Speaking	0.028	0.03	0.119	0.04	0.034	0.204**
Writing	0.147	0.03	Listening	0.105	0.03	0.042	0.05	−0.047	0.131
Listening	0.105	0.03	Speaking	0.028	0.03	0.077	0.04	−0.005	0.159
8th grade NECAP reading									
Reading	0.283	0.04	Writing	0.261	0.03	0.022	0.05	−0.079	0.123
Reading	0.283	0.04	Speaking	0.097	0.03	0.186	0.05	0.086	0.286**
Reading	0.283	0.04	Listening	0.018	0.04	0.265	0.06	0.156	0.374**
Writing	0.261	0.03	Speaking	0.097	0.03	0.164	0.05	0.071	0.257**
Writing	0.261	0.03	Listening	0.018	0.04	0.243	0.05	0.140	0.346**
Listening	0.018	0.04	Speaking	0.097	0.03	−0.079	0.05	−0.181	0.023
8th grade NECAP writing									
Reading	0.187	0.04	Writing	0.268	0.03	−0.081	0.05	−0.180	0.018
Reading	0.187	0.04	Speaking	0.135	0.03	0.052	0.05	−0.045	0.149
Reading	0.187	0.04	Listening	0.110	0.04	0.077	0.05	−0.030	0.184
Writing	0.268	0.03	Speaking	0.135	0.03	0.133	0.05	0.043	0.223**
Writing	0.268	0.03	Listening	0.110	0.04	0.158	0.05	0.058	0.258**
Listening	0.110	0.04	Speaking	0.135	0.03	−0.025	0.05	−0.124	0.074

(CONTINUED)

TABLE E1 (CONTINUED)

0.95 confidence interval around regression coefficients, by grade level and NECAP content area (within models), 2006

ACCESS subdomain	β	Standard error	ACCESS subdomain	β	Standard error	Difference between βs	Standard error of difference between βs	.95 confidence interval[a] Lower boundary	.95 confidence interval[a] Upper boundary
8th grade NECAP mathematics									
Reading	0.260	0.04	Writing	0.223	0.04	0.037	0.06	−0.077	0.151
Reading	0.260	0.04	Speaking	0.015	0.04	0.245	0.06	0.131	0.359**
Reading	0.260	0.04	Listening	0.001	0.05	0.259	0.06	0.134	0.384**
Writing	0.223	0.04	Speaking	0.015	0.04	0.208	0.05	0.103	0.313**
Writing	0.223	0.04	Listening	0.001	0.05	0.222	0.06	0.105	0.339**
Listening	0.001	0.05	Speaking	0.015	0.04	−0.014	0.06	−0.131	0.103

** .95 confidence interval does not contain 0.

a. The significance of the difference between the standardized regression coefficients was calculated by constructing a .95 confidence interval around the difference between the standardized regression coefficients. The interval was calculated as follows: $\beta_1 - \beta_2 \pm 1.96(SE_{\beta_1-\beta_2})$ where $SE_{\beta_1-\beta_2} = \sqrt{(SE_1)^2 + (SE_2)^2}$.

Source: Authors' calculations based on student English language learner scores from ACCESS for ELLs™ FAQ-test administration (2005) and student English language learner scores from the NECAP assessment from Measured Progress (2006).

TABLE E2

0.95 confidence interval around regression coefficients, by content areas (across 5th and 8th grade models), 2006

ACCESS subdomain	5th grade β	5th grade Standard error	8th grade β	8th grade Standard error	Difference between βs	Standard error of difference between βs	.95 confidence interval[a] Lower boundary	.95 confidence interval[a] Upper boundary
NECAP reading								
Reading	0.383	0.03	0.283	0.04	0.100	0.05	0.004	0.196**
Writing	0.239	0.03	0.261	0.03	−0.022	0.05	−0.111	0.067
Listening	0.031	0.03	0.018	0.04	0.013	0.05	−0.083	0.109
Speaking	0.093	0.03	0.097	0.03	−0.004	0.04	−0.085	0.077
NECAP writing								
Reading	0.303	0.03	0.187	0.04	0.116	0.05	0.020	0.212**
Writing	0.294	0.03	0.268	0.03	0.026	0.05	−0.063	0.115
Listening	0.048	0.03	0.110	0.04	−0.062	0.05	−0.157	0.033
Speaking	0.015	0.03	0.135	0.03	−0.120	0.04	−0.202	−0.038**
NECAP mathematics								
Reading	0.320	0.03	0.260	0.04	0.060	0.06	−0.048	0.168
Writing	0.147	0.03	0.223	0.04	−0.076	0.05	−0.175	0.023
Listening	0.105	0.03	0.001	0.05	0.104	0.06	−0.005	0.213
Speaking	0.028	0.03	0.015	0.04	0.013	0.05	−0.080	0.106

** .95 confidence interval does not contain 0.

a. The significance of the difference between the standardized regression coefficients was calculated by constructing a .95 confidence interval around the difference between the standardized regression coefficients. The interval was calculated as follows: $\beta_1 - \beta_2 \pm 1.96(SE_{\beta_1-\beta_2})$ where $SE_{\beta_1-\beta_2} = \sqrt{(SE_1)^2 + (SE_2)^2}$.

Source: Authors' calculations based on student English language learner scores from ACCESS for ELLs™ FAQ-test administration (2005) and student English language learner scores from the NECAP assessment from Measured Progress (2006).

APPENDIX F
MULTILEVEL MODELING PROCEDURES

Hierarchical linear regression models were formulated to examine the relationship between ACCESS domain scores and NECAP reading, writing, and mathematics scores while holding student and school covariates constant. Multilevel modeling procedures were used because they correctly model the dependence among individuals in the same school (that is, they produce unbiased estimates of the standard errors associated with the regression coefficients) and allow individual and group characteristics to be included simultaneously when modeling individual outcomes. Subsequent to running an unconditional model that included only a random school effect, three two-level regression models were formulated: model 1 included only student covariates, model 2 included student and school covariates, and model 3 included the student and school covariates and students' ACCESS domain scores in listening, speaking, reading, and writing. Model 3 allowed the research team to address the primary research question for this project:

> *How does performance in four language domains on an English language proficiency exam predict English language learner students' performance on a state content assessment after accounting for student and school characteristics?*

Each model used in the analysis was a two-level model in which English language learner students were nested within schools. The general two-level model assumes a random sample of *i* English language learner students within *j* schools, such that Y_{ij} is the outcome variable for English language learner student *i* in school *j* (Raudenbush and Bryk 2002). The general level-one or student model was:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \ldots + \beta_{kj}X_{kij} + r_{ij}$$

The NECAP student outcome variable for English language learner students (reading, writing, or mathematics scores), Y_{ij}, was modeled as a function of an intercept and a linear combination of student characteristics, X_{kij}. These X_{kij}s were student covariates only in models 1 and 2, and student covariates and ACCESS domain scores in listening, speaking, reading, and writing in model 3. The predicted outcome is composed of a unique intercept, β_{0j}, and slope for each predictor variable, β_{kj}, as well as a random student effect, r_{ij}.

Through empirical examination of the variability in the level-one regression coefficients across schools, the research team found no significant variation in the relationships between the level-one predictors (student covariates and ACCESS domain scores) and the NECAP outcome measures across schools. Therefore, the level-one slopes were fixed; the X_{kij} were constrained to have the same fixed value for each school. In this way only the level-one intercept was allowed to vary across schools. The general level-two or school models were:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \ldots + \gamma_{0P}W_{Pj} + u_{0j}$$
$$\beta_{kj} = \gamma_{k0} \text{ for } k = 1, 2, \ldots, k$$

For models 2 and 3 the variation in the level-one intercept across schools was modeled at the second level as a function of an intercept, γ_{00}, and a linear combination of school covariates, W_{pj}. Each school had a unique random effect, u_{0j}.

Model 3, the final intercept-only model that allowed the research team to examine the relationship between ACCESS scores and NECAP reading, writing, and mathematics scores while holding student and school covariates constant was:

Level one:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{student gender})_{ij} + \beta_{2j}\,(\text{student poverty status})_{ij} + \beta_{3j}\,(\text{student disability status})_{ij} + \beta_{4j}\,(\text{age status})_{ij} + \beta_{5j}\,(\text{years in English language learner program})_{ij} + \beta_{6j}\,(\text{Asian dummy variable})_{ij} + \beta_{7j}\,(\text{non-Hispanic Black dummy variable})_{ij} + \beta_{8j}\,(\text{Hispanic dummy variable})_{ij} + \beta_{9j}\,(\text{ACCESS listening subscore})_{ij} + \beta_{10j}\,(\text{ACCESS speaking subscore})_{ij} + \beta_{11j}\,(\text{ACCESS reading subscore})_{ij} + \beta_{12j}\,(\text{ACCESS writing subscore})_{ij} + r_{ij}$$

Level two:

$\beta_{0j} = \gamma_{00} + \gamma_{01}$ (school size)$_j$ + γ_{02} (school poverty proxy: percent receiving free or reduced-price lunch)$_j$ + γ_{03} (racial composition: percent White)$_j$ + γ_{04} (English language learner student density)$_j$ + γ_{05} (rural dummy variable)$_j$ + γ_{06} (urban dummy variable)$_j$ + γ_{07} (New Hampshire dummy variable)$_j$ + γ_{08} (Vermont dummy variable)$_j$ + u_{0j}

$\beta_{(1\rightarrow12)j} = \gamma_{(1\rightarrow12)0}$
for each level-one slope k = 1–12.

As described in appendix C, each NECAP outcome variable (Y_{ij}) was standardized to have a mean of 0 and a standard deviation of 1. The dichotomous level-one student covariates (gender, poverty status, disability status, age status, and the race dummy variables) were included in the model uncentered, and the only continuous student covariate (years in English language learner program) was centered around the grand mean. The ACCESS domain scores were standardized to have a mean of 0 and a standard deviation of 1 across all schools. Through standardizing the NECAP outcome scores and the ACCESS domain scores the regression coefficients could be compared across the reading, writing, and mathematics domains within a grade. The same students were included in the models for reading, writing, and mathematics at each grade level, and because they had the same background characteristics (such as gender, race, and the like) in each model, the regression coefficients for the student covariates were comparable across domains within the same grade.

At level two the continuous school covariates (school size, school poverty, school racial composition, and school English language learner student density) were rescaled and entered into the model grand mean centered. The dichotomous school covariates (rural, suburban, and urban dummy variables, and state location) were entered into the level-two models uncentered. For a description of how the NECAP outcome scores and the level one and level two covariates and ACCESS scores were coded and rescaled, see table C5 in appendix C.

Because only three states' data were included in the analyses, it was impossible to model state membership at a third level. For this reason the between-school variability will be confounded with the between-state variability in the models presented in this report.

In addition to the regression coefficients and their associated significance levels, the regression models allowed the research team to estimate the total percentage of variance in the NECAP outcome measures that was explained by the student and school covariates and the ACCESS English language domain scores. This percentage was calculated for each of the three models by comparing the residual variance to the available variance in the unconditional model. Specifically, the percentage of variance explained (analogous to R^2) was estimated for each model using the following equation:

$$1 - \frac{\text{Total residual variance under the conditional model}}{\text{Total unconditional variance}}$$

The results from the multilevel regression models are presented in tables G1–G6 of appendix G.

TABLE G1

Predictors of 5th grade NECAP reading scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Signifi-cance	Coefficient	Standard error	Signifi-cance	Coefficient	Standard error	Signifi-cance	Coefficient	Standard error	Signifi-cance
Intercept	0.131**	0.043	0.003	0.516**	0.075	< .001	0.492**	0.110	< .001	0.271**	0.089	0.003
Student model												
ACCESS predictors												
Listening										0.031	0.028	0.273
Speaking										0.093**	0.025	< .001
Reading										0.383**	0.030	< .001
Writing										0.239**	0.030	< .001
Student covariates												
Gender[a]				0.109**	0.049	0.026	0.106**	0.049	0.030	0.020	0.039	0.609
Poverty status[b]				−0.221**	0.062	< .001	−0.164**	0.064	0.011	−0.146**	0.051	0.004
Disability status[c]				−0.717**	0.071	< .001	−0.708**	0.072	< .001	−0.337**	0.058	< .001
Age status[d]				−0.234**	0.087	0.007	−0.240**	0.087	0.006	−0.079	0.069	0.254
Asian				0.008	0.095	0.931	0.010	0.095	0.920	−0.032	0.075	0.674
Non-Hispanic Black				−0.711**	0.110	< .001	−0.652**	0.112	< .001	−0.298**	0.090	0.001
Hispanic				−0.277**	0.082	0.001	−0.196**	0.089	0.028	−0.130	0.070	0.066
Years in English language learner programs				0.060**	0.016	< .001	0.064**	0.016	< .001	−0.061**	0.014	< .001
School model												
School covariates												
School size[e]							0.010	0.020	0.605	0.013	0.016	0.441
School poverty[f]							−0.088**	0.030	0.004	−0.005	0.025	0.852
Racial composition[g]							−0.049	0.033	0.142	−0.006	0.027	0.839
English language learner student density[h]							−0.027	0.037	0.473	0.017	0.031	0.590
Rural							0.219	0.121	0.071	0.221**	0.098	0.025
Urban							−0.097	0.103	0.349	−0.024	0.085	0.777

(CONTINUED)

TABLE G1 (CONTINUED)

Predictors of 5th grade NECAP reading scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Signifi-cance	Coefficient	Standard error	Signifi-cance	Coefficient	Standard error	Signifi-cance	Coefficient	Standard error	Signifi-cance
New Hampshire							−0.251**	0.124	0.044	−0.066	0.101	0.518
Vermont							−0.083	0.146	0.568	0.011	0.119	0.926
	Available variance		Signifi-cance	Residual variance	Signifi-cance of residual variance	Variance explained	Residual variance	Signifi-cance of residual variance	Variance explained	Residual variance	Signifi-cance of residual variance	Variance explained
Within schools	0.864	84.54%	< .001	0.755	< .001		0.752	< .001		0.462	< .001	
Between schools	0.158	15.46%	< .001	0.097	< .001		0.079	0.001		0.059	0.001	
Total	1.022			0.852		17%	0.831		19%	0.521		49%

** Statistically significant at $p < 0.05$.

a. Defined as 1 = female, 0 = male.

b. Defined as 1 = in poverty (eligible for free or reduced-price lunch), 0 = not in poverty.

c. Defined as 1 = with disabilities (has an Individualized Education Program), 0 = without disabilities.

d. Age status was 1 = overage, 0 = not overage.

e. Measured in units of 100 students.

f. Defined as the percent of students within the school eligible for free or reduced-price lunch. Measured in units of 10 percentage points.

g. Defined as the percent of students in the school who are White. Measured in units of 10 percentage points.

h. Defined as the percent of students in the school who are English language learners. Measured in units of 10 percentage points.

Note: Changes in NECAP scores are in standard deviation units. ACCESS predictors are measured in standard deviation units. Along with male, not in poverty, without disabilities, and not overage, omitted categories were non-Hispanic White, suburban schools, and Rhode Island schools.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

TABLE G2

Predictors of 5th grade NECAP writing scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
Intercept	0.148**	0.047	0.002	0.359**	0.077	< .001	0.385**	0.120	0.002	0.171	0.097	0.081
Student model												
ACCESS predictors												
Listening										0.048	0.029	0.103
Speaking										0.015	0.027	0.580
Reading										0.303**	0.031	< .001
Writing										0.294**	0.031	< .001
Student covariates												
Gender[a]				0.266**	0.048	< .001	0.265**	0.048	0.000	0.175**	0.041	< .001
Poverty status[b]				−0.146**	0.063	0.020	−0.104	0.064	0.108	−0.074	0.053	0.164
Disability status[c]				−0.641**	0.071	< .001	−0.646**	0.071	< .001	−0.312**	0.061	< .001
Age status[d]				−0.276**	0.086	0.001	−0.275**	0.086	0.001	−0.144	0.072	0.046
Asian				0.058	0.096	0.547	0.060	0.096	0.531	0.009	0.079	0.910
Non-Hispanic Black				−0.537**	0.111	< .001	−0.487**	0.112	< .001	−0.163	0.094	0.085
Hispanic				−0.238**	0.084	0.005	−0.171	0.090	0.057	−0.100	0.074	0.177
Years in English language learner programs				0.057**	0.016	< .001	0.063**	0.016	< .001	−0.047**	0.015	0.001
School model												
School covariates												
School size[e]							0.006	0.023	0.807	0.010	0.018	0.580
School poverty[f]							−0.062	0.034	0.072	0.013	0.028	0.635
Racial composition[g]							−0.026	0.038	0.506	0.020	0.031	0.511
English language learner student density[h]							−0.062	0.046	0.182	−0.015	0.036	0.685
Rural							0.102	0.132	0.440	0.112	0.107	0.296
Urban							−0.113	0.119	0.343	−0.038	0.095	0.688

(CONTINUED)

TABLE G2 (CONTINUED)

Predictors of 5th grade NECAP writing scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
New Hampshire							−0.332**	0.138	0.017	−0.160	0.112	0.152
Vermont							−0.262	0.162	0.107	−0.175	0.131	0.181
	Available variance		Significance	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained
Within schools	0.827	79.52%	< .001	0.721	< .001		0.721	< .001		0.500	< .001	
Between schools	0.213	20.48%	< .001	0.172	< .001		0.149	< .001		0.087	< .001	
Total	1.040			0.893		14%	0.870		16%	0.587		44%

** Statistically significant at $p < 0.05$.

a. Defined as 1 = female, 0 = male.

b. Defined as 1 = in poverty (eligible for free or reduced-price lunch), 0 = not in poverty.

c. Defined as 1 = with disabilities (has an Individualized Education Program), 0 = without disabilities.

d. Age status was 1 = overage, 0 = not overage.

e. Measured in units of 100 students.

f. Defined as the percent of students within the school eligible for free or reduced-price lunch. Measured in units of 10 percentage points.

g. Defined as the percent of students in the school who are White. Measured in units of 10 percentage points.

h. Defined as the percent of students in the school who are English language learners. Measured in units of 10 percentage points.

Note: Changes in NECAP scores are in standard deviation units. ACCESS predictors are measured in standard deviation units. Along with male, not in poverty, without disabilities, and not overage, omitted categories were non-Hispanic White, suburban schools, and Rhode Island schools.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

TABLE G3

Predictors of 5th grade NECAP mathematics scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
Intercept	0.128**	0.042	0.003	0.594**	0.075	< .001	0.549**	0.108	< .001	0.373**	0.093	< .001
Student model												
ACCESS predictors												
Listening										0.105**	0.031	0.001
Speaking										0.028	0.028	0.319
Reading										0.320**	0.033	< .001
Writing										0.147**	0.033	< .001
Student covariates												
Gender[a]				−0.063	0.050	0.204	−0.068	0.050	0.169	−0.132**	0.044	0.002
Poverty status[b]				−0.198**	0.063	0.002	−0.129**	0.065	0.047	−0.109	0.056	0.053
Disability status[c]				−0.680**	0.072	< .001	−0.669**	0.073	< .001	−0.384**	0.065	< .001
Age status[d]				−0.125	0.089	0.159	−0.136	0.089	0.126	−0.015	0.077	0.846
Asian				0.051	0.095	0.596	0.013	0.096	0.894	−0.026	0.083	0.752
Non-Hispanic Black				−0.708**	0.111	< .001	−0.323**	0.090	< .001	−0.395**	0.099	< .001
Hispanic				−0.343**	0.082	< .001	−0.689**	0.113	< .001	−0.275**	0.078	< .001
Years in English language learner programs				0.024	0.016	0.145	0.027	0.016	0.100	−0.072**	0.015	< .001
School model												
School covariates												
School size[e]							0.002	0.019	0.937	0.001	0.016	0.932
School poverty[f]							−0.117**	0.029	< .001	−0.050	0.025	0.048
Racial composition[g]							−0.118**	0.032	< .001	−0.078	0.027	0.005
English language learner student density[h]							−0.039	0.035	0.273	−0.004	0.029	0.897
Rural							0.131	0.119	0.273	0.109	0.102	0.290
Urban							−0.115	0.099	0.248	−0.041	0.084	0.626

(CONTINUED)

TABLE G3 (CONTINUED)

Predictors of 5th grade NECAP mathematics scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
New Hampshire							0.046	0.120	0.701	0.174	0.104	0.096
Vermont							0.183	0.142	0.200	0.229	0.123	0.063
	Available variance		Significance	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained
Within schools	0.875	87.06%	< .001	0.784	< .001		0.781	< .001		0.591	< .001	
Between schools	0.130	12.94%	< .001	0.076	< .001		0.061	0.001		0.041	0.01	
Total	1.005			0.860		14%	0.842		16%	0.632		37%

** Statistically significant at $p < 0.05$.

a. Defined as 1 = female, 0 = male.

b. Defined as 1 = in poverty (eligible for free or reduced-price lunch), 0 = not in poverty.

c. Defined as 1 = with disabilities (has an Individualized Education Program), 0 = without disabilities.

d. Age status was 1 = overage, 0 = not overage.

e. Measured in units of 100 students.

f. Defined as the percent of students within the school eligible for free or reduced-price lunch. Measured in units of 10 percentage points.

g. Defined as the percent of students in the school who are White. Measured in units of 10 percentage points.

h. Defined as the percent of students in the school who are English language learners. Measured in units of 10 percentage points.

Note: Changes in NECAP scores are in standard deviation units. ACCESS predictors are measured in standard deviation units. Along with male, not in poverty, without disabilities, and not overage, omitted categories were non-Hispanic White, suburban schools, and Rhode Island schools.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

TABLE G4

Predictors of 8th grade NECAP reading scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
Intercept	0.265**	0.069	< .001	0.495**	0.098	< .001	0.406**	0.135	0.004	0.234**	0.103	0.026
Student model												
ACCESS predictors												
Listening										0.018	0.040	0.661
Speaking										0.097**	0.033	0.004
Reading										0.283**	0.039	< .001
Writing										0.261**	0.034	< .001
Student covariates												
Gender[a]				0.125**	0.058	0.031	0.126**	0.057	0.029	0.063	0.049	0.196
Poverty status[b]				−0.125**	0.063	0.047	−0.117	0.064	0.066	−0.036	0.053	0.495
Disability status[c]				−0.844**	0.082	< .001	−0.840**	0.082	< .001	−0.487**	0.071	< .001
Age status[d]				−0.248**	0.083	0.003	−0.244**	0.083	0.003	−0.100	0.070	0.156
Asian				0.046	0.115	0.688	0.106	0.112	0.342	0.077	0.092	0.400
Non-Hispanic Black				−0.217	0.122	0.075	−0.119	0.122	0.329	−0.019	0.102	0.852
Hispanic				−0.204**	0.097	0.035	−0.138	0.098	0.159	−0.134	0.081	0.100
Years in English language learner programs				0.058**	0.013	< .001	0.051**	0.013	< .001	−0.007	0.012	0.585
School model												
School covariates												
School size[e]							−0.014	0.018	0.447	−0.003	0.013	0.826
School poverty[f]							−0.071	0.044	0.115	−0.051	0.033	0.133
Racial composition[g]							−0.044	0.038	0.255	−0.042	0.027	0.133
English language learner student density[h]							−0.091**	0.027	0.002	−0.034	0.020	0.102
Rural							0.126	0.153	0.409	0.161	0.119	0.178
Urban							−0.170	0.145	0.248	−0.087	0.104	0.413

(CONTINUED)

TABLE G4 (CONTINUED)

Predictors of 8th grade NECAP reading scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
New Hampshire							–0.169	0.159	0.289	–0.114	0.120	0.346
Vermont							0.121	0.181	0.505	0.114	0.140	0.416
	Available variance		Signifi-cance	Residual variance	Signifi-cance of residual variance	Variance explained	Residual variance	Signifi-cance of residual variance	Variance explained	Residual variance	Signifi-cance of residual variance	Variance explained
Within schools	0.764	73.10%	< .001	0.669	< .001		0.676	< .001		0.480	< .001	
Between schools	0.281	26.90%	< .001	0.173	< .001		0.077	< .001		0.029	< .001	
Total	1.045			0.842		19%	0.754		28%	0.509		51%

** Statistically significant at $p < 0.05$.

a. Defined as 1 = female, 0 = male.

b. Defined as 1 = in poverty (eligible for free or reduced-price lunch), 0 = not in poverty.

c. Defined as 1 = with disabilities (has an Individualized Education Program), 0 = without disabilities.

d. Age status was 1 = overage, 0 = not overage.

e. Measured in units of 100 students.

f. Defined as the percent of students within the school eligible for free or reduced-price lunch. Measured in units of 10 percentage points.

g. Defined as the percent of students in the school who are White. Measured in units of 10 percentage points.

h. Defined as the percent of students in the school who are English language learners. Measured in units of 10 percentage points.

Note: Changes in NECAP scores are in standard deviation units. ACCESS predictors are measured in standard deviation units. Along with male, not in poverty, without disabilities, and not overage, omitted categories were non-Hispanic White, suburban schools, and Rhode Island schools.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

TABLE G5

Predictors of 8th grade NECAP writing scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
Intercept	0.240**	0.068	0.001	0.454**	0.097	< .001	0.473**	0.145	0.002	0.289**	0.116	0.015
Student model												
ACCESS predictors												
Listening										0.110**	0.039	0.005
Speaking										0.135**	0.032	< .001
Reading										0.187**	0.038	< .001
Writing										0.268**	0.033	< .001
Student covariates												
Gender[a]				0.251**	0.056	< .001	0.253**	0.056	< .001	0.199**	0.047	< .001
Poverty status[b]				−0.097	0.062	0.116	−0.088	0.062	0.158	0.000	0.052	0.995
Disability status[c]				−0.680**	0.080	< .001	−0.677**	0.080	< .001	−0.337**	0.069	< .001
Age status[d]				−0.206**	0.082	0.012	−0.198**	0.082	0.016	−0.046	0.068	0.496
Asian				−0.100	0.113	0.374	−0.064	0.111	0.565	−0.076	0.091	0.404
Non-Hispanic Black				−0.279**	0.120	0.020	−0.209	0.120	0.081	−0.103	0.099	0.297
Hispanic				−0.316**	0.095	0.001	−0.279**	0.097	0.004	−0.272**	0.079	0.001
Years in English language learner programs				0.068**	0.013	< .001	0.062**	0.013	< .001	−0.006	0.012	0.605
School model												
School covariates												
School size[e]							−0.022	0.019	0.266	−0.010	0.016	0.531
School poverty[f]							−0.041	0.048	0.394	−0.014	0.040	0.732
Racial composition[g]							−0.047	0.042	0.267	−0.046	0.035	0.190
English language learner student density[h]							−0.094**	0.030	0.003	−0.041	0.026	0.119
Rural							−0.023	0.161	0.884	−0.003	0.135	0.980
Urban							−0.258	0.161	0.114	−0.169	0.134	0.215

(CONTINUED)

TABLE G5 (CONTINUED)

Predictors of 8th grade NECAP writing scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
New Hampshire							–0.130	0.170	0.445	–0.054	0.142	0.703
Vermont							0.172	0.193	0.375	0.213	0.162	0.193
	Available variance		Significance	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained
Within schools	0.726	71.80%	< .001	0.641	< .001		0.645	< .001		0.436	< .001	
Between schools	0.281	28.20%	< .001	0.181	< .001		0.115	< .001		0.068	< .001	
Total	1.007			0.821		18%	0.760		25%	0.504		50%

** Statistically significant at $p < 0.05$.

a. Defined as 1 = female, 0 = male.

b. Defined as 1 = in poverty (eligible for free or reduced-price lunch), 0 = not in poverty.

c. Defined as 1 = with disabilities (has an Individualized Education Program), 0 = without disabilities.

d. Age status was 1 = overage, 0 = not overage.

e. Measured in units of 100 students.

f. Defined as the percent of students within the school eligible for free or reduced-price lunch. Measured in units of 10 percentage points.

g. Defined as the percent of students in the school who are White. Measured in units of 10 percentage points.

h. Defined as the percent of students in the school who are English language learners. Measured in units of 10 percentage points.

Note: Changes in NECAP scores are in standard deviation units. ACCESS predictors are measured in standard deviation units. Along with male, not in poverty, without disabilities, and not overage, omitted categories were non-Hispanic White, suburban schools, and Rhode Island schools.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

TABLE G6

Predictors of 8th grade NECAP mathematics scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
Intercept	0.224**	0.061	< .001	0.498**	0.097	< .001	0.278**	0.120	0.022	0.137	0.104	0.190
Student model												
ACCESS predictors												
Listening										0.001	0.046	0.977
Speaking										0.015	0.038	0.689
Reading										0.260**	0.044	< .001
Writing										0.223**	0.038	< .001
Student covariates												
Gender[a]				−0.057	0.061	0.354	−0.053	0.061	0.386	−0.110**	0.056	0.048
Poverty status[b]				−0.071	0.066	0.281	−0.044	0.067	0.505	0.019	0.061	0.749
Disability status[c]				−0.813**	0.086	< .001	−0.794**	0.086	< .001	−0.497**	0.082	< .001
Age status[d]				−0.129	0.088	0.145	−0.133	0.088	0.129	−0.024	0.081	0.765
Asian				0.195	0.118	0.099	0.277	0.114	0.015	0.228**	0.102	0.026
Non-Hispanic Black				−0.325**	0.127	0.011	−0.196	0.127	0.124	−0.115	0.116	0.320
Hispanic				−0.290**	0.099	0.003	−0.184	0.102	0.070	−0.182**	0.092	0.049
Years in English language learner programs				0.043**	0.014	0.002	0.032**	0.014	0.022	−0.005	0.013	0.685
School model												
School covariates												
School size[e]							−0.016	0.015	0.304	−0.007	0.013	0.584
School poverty[f]							−0.117**	0.038	0.003	−0.115**	0.032	0.001
Racial composition[g]							−0.078**	0.030	0.014	−0.087**	0.024	0.001
English language learner student density[h]							−0.058**	0.022	0.012	−0.014	0.019	0.477
Rural							0.155	0.141	0.274	0.197	0.124	0.113
Urban							−0.044	0.113	0.699	0.033	0.090	0.720

(CONTINUED)

TABLE G6 (CONTINUED)

Predictors of 8th grade NECAP mathematics scores, 2006

	Unconditional model			Model 1			Model 2			Model 3		
Predictor	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance	Coefficient	Standard error	Significance
New Hampshire							0.007	0.138	0.960	0.055	0.116	0.638
Vermont							0.264	0.162	0.107	0.240	0.141	0.091
	Available variance		Significance	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained	Residual variance	Significance of residual variance	Variance explained
Within schools	0.855	83.60%	< .001	0.769	< .001		0.772	< .001		0.645	< .001	
Between schools	0.168	16.40%	< .001	0.079	< .001		0.021	< .001		0.007	< .001	
Total	1.023			0.848		17%	0.794		22%	0.652		36%

** Statistically significant at $p < 0.05$.

a. Defined as 1 = female, 0 = male.

b. Defined as 1 = in poverty (eligible for free or reduced-price lunch), 0 = not in poverty.

c. Defined as 1 = with disabilities (has an Individualized Education Program), 0 = without disabilities.

d. Age status was 1 = overage, 0 = not overage.

e. Measured in units of 100 students.

f. Defined as the percent of students within the school eligible for free or reduced-price lunch. Measured in units of 10 percentage points.

g. Defined as the percent of students in the school who are White. Measured in units of 10 percentage points.

h. Defined as the percent of students in the school who are English language learners. Measured in units of 10 percentage points.

Note: Changes in NECAP scores are in standard deviation units. ACCESS predictors are measured in standard deviation units. Along with male, not in poverty, without disabilities, and not overage, omitted categories were non-Hispanic White, suburban schools, and Rhode Island schools.

Source: Authors' calculations based on student English language learner scores and demographic data from ACCESS for ELLs™ FAQ-test administration (2005), student English language learner scores and demographic data from the NECAP assessment from Measured Progress (2006), and school data from U.S. Department of Education, National Center for Education Statistics (2007).

NOTES

1. In 2007/08 the NECAP also began to assess students in science.

2. The ACCESS assessment is administered from January through February in New Hampshire and Rhode Island and from March through April in Vermont. The NECAP assessments in reading, writing, and mathematics are administered in October in all three states. English language learner students in 4th and 7th grades who took the ACCESS assessments in early or mid-spring 2006 are expected to have taken the 5th and 8th grade NECAP content assessments, respectively, in fall 2006.

3. Whereas the NECAP reading and mathematics assessments are administered each year to all students in grades 3–8, the NECAP writing assessment is administered each year to students in grades 5 and 8 only.

4. This score represents the intercept in model 3 for each NECAP outcome. In other words, the score is the predicted NECAP score for the English language learner student whose ACCESS domain scores and background characteristics equal the sample grand mean or zero. Based on the sample data and the definition of covariates, this student achieved the average score for the entire 5th or 8th grade sample in each ACCESS domain. This student was also male, White, not in poverty, and without disabilities; he spent an average number of years in English language learner programs and attended a Rhode Island suburban school of average size, poverty level, racial distribution, and English language learner student density.

5. See appendix E for significance tests examining differences between regression coefficients.

6. In 2006 the ACCESS scores of academic language in content areas included too few items for analysis.

7. For example, Cummins (1981a) described a distinction between the acquisition of language required for social interactions and that required for academic communication. Language learners typically demonstrate proficiency communicating orally in context-rich situations (playground, grocery store) before they can achieve proficiency in the more formal, context-independent academic language that is the medium for most classroom learning and assessments.

8. Predicted NECAP score changes measured in standard deviation units are from a one standard deviation unit increase in ACCESS scores, and predicted NECAP score changes measured in scale score points are from a 10 point increase in ACCESS scale scores.

9. For example, 10 scale score points represent over a third of a standard deviation in 5th grade ACCESS writing scores. In contrast, 10 scale score points represent a seventh of a standard deviation in 5th grade ACCESS speaking scores (see table C6 in appendix C).

10. The significance of the difference between the standardized regression coefficients was calculated by constructing a .95 confidence interval around the difference between the standardized regression coefficients. The interval was calculated as follows: $\beta_1 - \beta_2 \pm 1.96(SE_{\beta_1-\beta_2})$ where $SE_{\beta_1-\beta_2} = \sqrt{(SE_1)^2 + (SE_2)^2}$. When 0 was not in the interval around the difference between the regression coefficients, it was concluded that there was a statistically significant difference between the coefficients. This means that over an infinite number of random samples, there is 95 percent confidence that the interval around the difference between the regression coefficients does not include the population mean difference between the regression coefficients.

11. As noted, due to the timing of ACCESS and NECAP testing, 4th grade English language learner students who took the ACCESS

assessment in spring 2006 should have taken the 5th grade NECAP assessment in fall 2006. Similarly, 7th grade English language learner students who took the ACCESS assessment in spring 2006 should have taken the 8th grade NECAP assessment in fall 2006.

12. Within the original NECAP database, some students were recorded as having taken the NECAP, but scores for all three content areas (reading, writing, and mathematics) were missing. Within the original ACCESS database, some student records were without ACCESS scores. In other cases student records did not have ACCESS scores in the four language domains (listening, speaking, reading, and writing) examined in this project, but had scores in English language proficiency for specific subject areas (such as mathematics and science), which were not examined in this report.

13. The sum of the number of cases with missing NECAP, ACCESS, and student or school background variables exceeds the total number of cases with any missing data because one student record in the 4th and 5th grade ACCESS-NECAP dataset was missing data for multiple variables. In the 7th and 8th grade dataset five student records were missing data for multiple variables.

14. Cited proportions of all students and schools from the three states were calculated from 2005/06 data from the National Center for Education Statistics, Common Core of Data.

15. NECAP scaled scores are three-digit scores, where the first digit indicates the grade level of the student and the second two digits are designed to range from 00 to 80. Students who receive a scaled score of 40 or above for their grade level (for example, 540 or 840) are designated as "proficient."

16. Each grade cluster has three tiers (from beginning to advanced English). Students take one of the three tiers. The reliability estimates for reading, writing, and listening are averaged across the three tiers (see Measured Progress 2006 for more information).

REFERENCES

Abedi, J., and Dietel, R. (2004). *Challenges in the No Child Left Behind Act for English language learners* (Policy Brief 7). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., and Lord, C. (2001). *NAEP math performance and test accommodations: interactions with student language background* (CSE Technical Report 536). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Leon, S., and Mirocha, J. (2003). *Impact of students' language background on content-based data: analyses of extant data* (CSE Technical Report 603). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abella, R., Urrutia, J., and Shneyderman, A. (2005). An examination of the validity of English-language achievement test scores in an English language learner population. *Bilingual Research Journal, 29*(1), 127–144.

ACCESS for ELLs™ FAQ-test administration. (2005). Retrieved December 3, 2005, from http://www.wida.us/ACCESSForELLs/faq_admin/.

Albus, D., Klein, J., Liu, K., and Thurlow, M. (2004). *Connecting English language proficiency, statewide assessments, and classroom proficiency* (LEP Projects Report 5). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

August, D., and Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bailey, A., and Butler, F. A. (2007). A conceptual framework of academic English language for broad application to education. In A. Bailey (Ed.), *The language demands of school: putting academic English to the test*. New Haven, CT: Yale University Press.

Bifulco, R., and Ladd, H. F. (2007). School choice, racial segregation and test-score gaps: evidence from North Carolina's charter school program. *Journal of Policy Analysis and Management, 26*(1), 31–56.

Brown, C. L. (2005). Equity of literacy-based math performance assessments for English language learners. *Bilingual Research Journal, 29*(2), 337–364.

Butler, F. A., Lord, C., Stevens, R. A., Borrego, M., and Bailey, A. (2004). *An approach to operationalizing academic language for language test development purposes: evidence from 5th-grade science and math.* Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Caldas, S. J. (1993). Reexamination of input and process factor effects in public school achievement. *Journal of Educational Research, 86*(4), 206–214.

Cazabon, M. T., Nicoladis, E., and Lambert, W. E. (January 1, 1998). *Becoming bilingual in the Amigos two-way immersion program* (Research Report 3). Berkeley, CA: Center for Research on Education, Diversity and Excellence. Retrieved September 15, 2007, from http://repositories.cdlib.org/crede/rsrchrpts/rr03.

Cole, N. (1997). *The ETS gender study: how females and males perform in educational settings*. Princeton, NJ: Educational Testing Service.

Coley, R. J. (2001). *Differences in the gender gap: comparisons across racial/ethnic groups in education and work*. Princeton, NJ: Educational Testing Service.

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly, 21*(4), 617.

Council of Chief State School Officers. (1992). *Summary of recommendations and policy implications for improving the assessment and monitoring of students with limited English proficiency.* Retrieved December 22, 2008, from http://www.ccsso.org/About_the_Council/policy_statements/1559.cfm

Cummins, J. (1981a). Age on arrival and immigrant second language learning in Canada: a reassessment. *Applied Linguistics, 2*, 132–149.

Cummins, J. (1981b). 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 status: a theoretical framework*. Los Angeles, CA: Evaluation, Dissemination and Assessment Center, California State University.

DeJong, E. J. (2002). Effective bilingual education: from theory to academic achievement in a two-way bilingual program. *Bilingual Research Journal, 26*(1), 1–20.

Fowler, W. J., and Walberg, H. J. (1991). School size, characteristics, and outcomes. *Educational Evaluation and Policy Analysis, 13*(2), 189–202.

Francis, D. J., Rivera, M., Lesaux, N., and Rivera, H. (2006). *Research-based recommendations for the use of accommodations in large-scale assessments*. Houston, TX: Center on Instruction.

Freeman, C. E. (2004). *Trends in educational equity of girls and women: 2004*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Gardner, V. A. (2001, October). *Does high school size matter for rural schools and students?* Paper presented at the meeting of the New England Educational Research Organization, Portsmouth, NH.

Geva, E. (2006). Second-language oral proficiency and second-language literacy. In D. August and T. Shanahan (Eds.), *Developing literacy in second-language learners: report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gottlieb, M. (2003). *Large-scale assessment of English language learners: addressing educational accountability in K-12 settings.* (Professional Paper 6). Alexandria, VA: Teachers of English to Speakers of Other Languages.

Gottlieb, M. (2004). *English language proficiency standards for English language learners in kindergarten through grade 12: frameworks for large-scale state and classroom assessment.* Madison, WI: Wisconsin Department of Public Instruction.

Gottlieb, M. (2006). *Assessing English language learners: bridges from language proficiency to academic achievement.* Thousand Oaks, CA: Corwin Press.

Hakuta, K. (2000). *How long does it take English learners to attain proficiency?* Palo Alto, CA: University of California Linguistic Minority Research Institute.

Kenyon, D. M., MacGregor, D., Louguit, M., Cho, B., and Ryu, J. R. (2007). *Annual technical report for ACCESS for ELLs® English language proficiency test, series 101, 2005–2006 administration.* Madison, WI: Center for Applied Linguistics, Board of Regents of the University of Wisconsin System on behalf of the WIDA Consortium.

Kids Count Data Center. (2006). *Children that speak a language other than English at home: 2006.* Annie E. Casey Foundation. Retrieved May 1, 2008, from http://www.kidscount.org/datacenter/compare_results.jsp?i=510.

Klecker, B. M. (2006). The gender gap in NAEP fourth-, eighth-, and twelfth-grade reading scores across years. *Reading Improvement, 43*(1), 50–56.

Kohler, A. D., and Lazarín, M. (2007). *Hispanic education in the United States* (NCLR Statistical Brief 8). Washington, DC: National Council of La Raza.

Lee, J., Grigg, W., and Dion, G. (2007a). *The nation's report card: mathematics 2007* (NCES 2007–494). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Lee, J., Grigg, W., and Donahue, P. (2007b). *The nation's report card: reading 2007* (NCES 2007–496). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: the case of school effects. *Educational Psychologist, 35*(2), 125–141.

Lindholm-Leary, K. J. (2001). *Dual language education.* Clevedon, UK: Multilingual Matters.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley.

Ma, X., and Wilkins, J. L. M. (2002). The development of science achievement in middle and high school: individual differences and school effects. *Evaluation Review, 26*(4), 23.

Mahoney, K. S., and MacSwan, J. (2005). Reexamining identification and reclassification of English language learners: A critical discussion of select state practices. *Bilingual Research Journal, 29*(1), 31–42.

Mahoney, K. S., Haladyna, T., and MacSwan, J. (2006, April). *A validity study of the SELP (Stanford English language proficiency) test as a tool for reclassifying ELLs.* Paper presented at the annual meeting of the American Education Research Association, San Francisco, CA.

McMillen, B. J. (2004). School size, achievement, and achievement gaps. *Education Policy Analysis Archives, 12*(58). Retrieved April 5, 2008 from http://epaa.asu.edu/epaa/v12n58/.

Meadows, S. O., Land, K. D., and Lamb, V. L. (2005). Assessing Gilligan vs. Sommers: gender-specific trends in child and youth well-being in the United States, 1985–2001. *Social Indicators Research, 70*, 1–52.

Measured Progress. (2006). *New England Common Assessment Program 2005–2006 technical report.* Dover, NH: Measured Progress.

Muijs, D., and Reynolds, D. (2003). Student background and teacher effects on achievement and attainment in mathematics: a longitudinal study. *Educational Research and Evaluation, 9*(3), 289–314.

Nowell, A., and Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: an analysis of differences in the mean, variance, and extreme scores. *Sex Roles, 39*(112), 21–42.

Rabinowitz, S., Ananda, S., and Bell, A. (2005). Strategies to assess the core academic knowledge of English language learners. *Journal of Applied Testing Technology, 7*(1), 1–12.

Ramírez, J. D. (1992). Executive summary of the final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual programs for language-minority students. *Bilingual Research Journal, 16*(1&2), 1-62.

Raudenbush, S. W., and Bryk, A. (2002). *Hierarchical linear models: applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Saunders, W. M., Foorman, B. R., and Carlson, C. D. (2006). Is a separate block of time for oral English language development in programs for English learners needed? *Elementary School Journal, 107*(2), 181–198.

Scarcella, R. (2003). *Academic English: a conceptual framework.* Irvine, CA: University of California Linguistic Minority Research Institute.

Solórzano, R. W. (2008). High stakes testing: issues, implications, and remedies for English language learners. *Review of Educational Research, 78*(2), 260–329.

Stevens, R. A., Butler, F. A., & Castellón-Wellington, M. (2001). *Academic language and content assessment: measuring the progress of English language learners* (CSE Technical Report 552). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Thomas, W. P., and Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. (2007). *Common core of data. Public elementary/secondary school universe survey 2005–06* [Data file]. Available from http://nces.ed.gov/ccd/.

Valdés, G. (2004). Between support and marginalisation: the development of academic language in linguistic minority children. *Bilingual Education and Bilingualism, 7*(2&3), 102–132.

Zehler, A. M., Hopstock, P. J., Fleischman, H. L., and Greniuk, C. (1994). *An examination of assessment of limited English proficient students.* Task Order D070 Report. Arlington, VA: Special Issues Analysis Center, Development Associates, Inc.