

**Research Report**  
ETS RR-14-23

# Improving Content Assessment for English Language Learners: Studies of the Linguistic Modification of Test Items

---

John W. Young

Teresa C. King

Maurice Cogan Hauck

Mitchell Ginsburgh

Lauren Kotloff

Julio Cabrera

Carlos Cavalie

December 2014

# ETS Research Report Series

---

## EIGNOR EXECUTIVE EDITOR

James Carlson  
*Principal Psychometrician*

## ASSOCIATE EDITORS

Beata Beigman Klebanov  
*Research Scientist*

Heather Buzick  
*Research Scientist*

Brent Bridgeman  
*Distinguished Presidential Appointee*

Keelan Evanini  
*Managing Research Scientist*

Marna Golub-Smith  
*Principal Psychometrician*

Shelby Haberman  
*Distinguished Presidential Appointee*

Donald Powers  
*Managing Principal Research Scientist*

Gautam Puhan  
*Senior Psychometrician*

John Sabatini  
*Managing Principal Research Scientist*

Matthias von Davier  
*Senior Research Director*

Rebecca Zwick  
*Distinguished Presidential Appointee*

## PRODUCTION EDITORS

Kim Fryer  
*Manager, Editing Services*

Ayleen Stellhorn  
*Editor*

---

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

## RESEARCH REPORT

# Improving Content Assessment for English Language Learners: Studies of the Linguistic Modification of Test Items

John W. Young, Teresa C. King, Maurice Cogan Hauck, Mitchell Ginsburgh, Lauren Kotloff, Julio Cabrera, & Carlos Cavalie

Educational Testing Service, Princeton, NJ

This article describes two research studies conducted on the linguistic modification of test items from K–12 content assessments. In the first study, 120 linguistically modified test items in mathematics and science taken by fourth and sixth graders were found to have a wide range of outcomes for English language learners (ELLs) and non-ELLs, with regard to performance on the original and modified versions of the same items. The original items were disclosed items from two states and were used in their standards-based assessments. The modified versions of these items were developed by a team of Educational Testing Service (ETS) assessment developers and researchers. The ETS research team identified several kinds of modifications that appeared promising and applied the modifications systematically. Ideally, modifying items leads to improved performance by ELLs while having little to no impact on the performance of non-ELLs. However, both groups of students performed better on some items, about the same on some other items, and worse on yet some other items. Cognitive interviews were conducted as a follow-up study to investigate which features of the linguistically modified items produced the observed outcomes for ELLs and non-ELLs. The results from this study added little clarity with regards to which of the linguistic modifications were effective and how they improved the understanding and performance of ELLs on these content items.

**Keywords** English language learners; cognitive interviews; linguistic modification; content assessments; K–12 students

doi:10.1002/ets2.12023

English language learners (ELLs) are one of the fastest growing subpopulations of students in U.S. schools (Kindler, 2002). By the year 2025, one projection estimates that one in four students nationally in grades K–12 will be an ELL (National Clearinghouse for English Language Acquisition, 2006). These demographic changes are of great importance for education in the United States. A major point of concern is that educational attainment for ELLs lags considerably behind that of non-ELLs. According to the 2007 National Assessment of Educational Progress (NAEP) test results, fourth grade ELLs scored an average of 36 points lower than non-ELLs in reading (Lee, Grigg, & Donahue, 2007) and 25 points lower in mathematics (Lee, Grigg, & Dion, 2007). In this article, we report on two studies in the continuing effort to better understand the factors that influence the performance of ELLs on content assessments. Identifying these factors will enable the development of more valid assessments and may help to reduce the achievement gap between ELLs and non-ELLs.

The concepts behind the linguistic modification of test items are consistent with those of universal test design. According to the No Child Left Behind Act of 2001, universally designed assessments are those that are “designed to be valid and accessible for use by the widest possible range of students, including students with disabilities and students with limited English proficiency” (No Child Left Behind Act, 2002). However, for items that were not designed using principles of universal test design, linguistic modification (sometimes referred to as the use of simplified or plain English) may be an appropriate testing accommodation for some students, including most ELLs. Studies have shown that the higher the level of linguistic complexity in a test item, the greater the performance gap between ELLs and non-ELLs, regardless of the difficulty of the content (Martiniello, 2008). Thus, language-based testing modifications may be more effective than other types of testing accommodations for ELLs.

In this article, we present the findings from two research studies. In the first study, we modified 120 disclosed, standards-based test items in mathematics and science. The original and modified versions of these items were administered to samples of fourth and sixth grade ELL and non-ELL students, and performance on both types of items, original and modified, were compared. Because the findings from the first study were mixed and not consistent with our hypotheses, we conducted a qualitative investigation using cognitive interviews as a follow-up study. In the second study, the main

*Corresponding author:* J. W. Young, E-mail: [jwyoung@ets.org](mailto:jwyoung@ets.org)

research question we addressed was: which features of the linguistically modified items produced the observed outcomes for ELLs and non-ELLs?

### Study 1: An Experimental Investigation of Linguistic Modification

This study investigated the impact of the linguistic modification of test items on the performance of ELLs and non-ELLs on academic content tests. The purpose of linguistic modification is to reduce or eliminate unnecessary linguistic complexity in test items and assessments. In subject-area content tests, high verbal processing demands may be differentially detrimental for the performance of ELLs. Linguistic modification lessens the verbal processing requirements for all examinees, while ensuring that measurement of the construct of interest is maintained. In this experimental study, we created test forms using disclosed state standards-based test items in mathematics and science in their original versions, as well as linguistically modified versions of the same items. ELLs and non-ELLs in fourth and sixth grades in two states were administered these test forms during the 2007–2008 and 2008–2009 school years.

In this study, the two research questions were:

1. Does the linguistic modification of mathematics and science test items given to fourth and sixth graders reduce the performance gap between ELLs and non-ELLs?
2. Are certain types of linguistic modification more effective than others in reducing the performance gap between ELLs and non-ELLs on mathematics and science tests?

### Prior Research

The linguistic modification of test items has been proposed as one testing accommodation that could be beneficial for ELLs. A review by Abedi, Lord, Hofstetter, and Baker (2000) found that linguistic modification was the only accommodation out of four that were studied that narrowed the score gap between ELLs and non-ELLs. Also, Abedi and Lord (2001) found that linguistically modifying eighth grade mathematics National Assessment of Educational Progress (NAEP) items produced small but significant changes in the scores of low and average ability students. The meta-analysis by Sireci, Li, and Scarpati (2003) concluded that there were mixed findings on linguistic modification, since some studies found that non-ELLs sometimes performed worse on the modified items, which contributed to a shrinking of the performance gap between ELLs and non-ELLs. A more recent review by Abedi (2007), which also included more recent studies, concluded that linguistic modification is a more valid and effective alternative to conventional testing approaches that do not account for the language-based needs of ELLs.

Two recent meta-analyses (Kieffer, Lesaux, Rivera, & Francis, 2009; Pennock-Roman & Rivera, 2011) found that the linguistic modification of test items was minimally effective as a testing accommodation for ELLs. Although the two meta-analyses included the same linguistic modification studies, the conclusions reached were substantially different. Kieffer et al. (2009) were more pessimistic in their evaluation of linguistic modification. They concluded that “... simplified English was not found to have a statistically significant effect” (p. 1182) and that “... there is little reason to be optimistic about the potential effectiveness of simplified English as a test accommodation” (p. 1187).

In contrast, Pennock-Roman and Rivera (2011) disagreed with these conclusions, since the focus on *average effect* masked the fact that several studies did find positive outcomes from the use of simplified English. In one study (Aguirre-Muñoz, 2000) that incorporated a measure of English proficiency into the research design, simplified English had a significant positive effect on the performance of ELLs classified as having high English proficiency. Additionally, the greater effectiveness of plain English for students with intermediate levels of English proficiency was found in two other studies. Pennock-Roman and Rivera concluded that judgments on the effectiveness of linguistic modification should be suspended until there are further investigations of this testing accommodation. In particular, additional studies should include a measure of students’ English proficiency level, as well as an evaluation of the linguistic complexity of the original and plain English versions of test items.

### Research Methods

#### Participants

School districts in several states with large numbers of ELLs were identified for participation. ELLs were identified based on the classification system used in each district—this was usually based on scores on the state’s English proficiency

test and teachers' recommendations. Testing administration was handled by a site coordinator at each school, usually the bilingual/ESL coordinator. The study design involved developing multiple test forms for mathematics and science for each grade. Each student was administered either a mathematics or a science test appropriate to his or her grade. For each grade and each subject, two versions of each test were created with a combination of linguistically modified and unmodified items, such that the two versions were counterbalanced. So, for example, if Item 1 on Form 1 of the mathematics test for grade 4 was an original standards-based test item, then Item 1 on Form 2 of the mathematics test for grade 4 would be the linguistically modified version of the same item. Thus, each student received only one version of each item—either the original or the modified. Test forms were systematically spiraled for distribution within each classroom to avoid any confounding of performance with test form that might be caused by improper distribution of the test forms.

### ***Instruments***

A large number of disclosed test items in mathematics and science were reviewed by the research team of researchers and assessment developers. Test items considered amenable to linguistic modification were selected and modified and then reviewed by two content experts to ensure that the item construct had not changed during linguistic modification. The team documented the types of modification that were made to each item to allow analysis by type of modification. For each subject and grade, a common set of NAEP items was also included to serve as a proxy for the ability levels of students responding to the different test forms. There was a total of 36 items for each of the eight test forms: six NAEP items, 15 original items, and 15 modified items. There were 120 experimental items across all forms, with each item appearing in its original and modified version (but on different forms).

### ***Linguistic Modification of Test Items***

The purpose of linguistic modification was to create test items that are more accessible to ELLs, while preserving the underlying construct. Instead of working with predetermined principles or rules for modifying test items, the research team made holistic changes to the items while not restricting the modifications to any particular type or category of changes. Some items went through multiple types of modifications, while some required just one type of modification to be made more accessible. The types of modification made to each item were noted. The modifications for each item were then reviewed and categories of modification were identified at this point.

This application of linguistic modification was independent of the applications of Language Muse (Burstein *et al.*, 2012) or other ETS-developed tools, as it derived from the expertise of assessment developers at ETS who have worked on Title 1 state testing programs with a specific emphasis on the use of plain English as a testing accommodation for ELLs. For this study, the team that developed the categories of linguistic modification consisted of a research scientist with expertise in second language acquisition and two assessment developers with expertise and substantial experience in the linguistic modification of test items from mathematics and science K–12 content assessments. This group's efforts were then further reviewed by the larger research team, which included additional research scientists with expertise in assessing ELLs for English proficiency and content knowledge.

The categories of linguistic modification were developed through an iterative process whereby the researchers started without any preconceived notions of what the categories should be. The categories were identified through a process by which each member of the research team independently classified items. The goal was to develop a valid and reliable classification scheme that reflected the main types of modifications employed. After multiple iterations, the team agreed on the 11 categories listed below; these categories were reviewed and accepted by the larger project team.

The 11 categories of linguistic modification that were developed and applied to the test items were:

1. **Remove Empty Context:** In cases where a context has been *tacked on* to a task, remove the context and set the task directly.
2. **Refine Context:** This refers to either making a context more explicit or recasting the item in a more accessible context.
3. **Simplify Vocabulary:** Identify challenging words not related to the content being assessed and replace them with more accessible words.
4. **Unpack:** Present complex ideas in a series of simple steps, following logical and/or chronological order as closely as possible.

**Table 1** Sample Sizes by Language Status and Grade Level

Grade level	English language learner	Non-English language learner	Total
Fourth	566	638	1,204
Sixth	366	490	856
Total	932	1,128	2,060

**Table 2** Comparisons of *p*-Values by Language Status for Grade 4 Mathematics

<i>p</i> -value comparison	English language learner	Non-English language learner
Original = Modified	12	11
Original > Modified	9	6
Original < Modified	9	13

5. **Make Item Stem Concise:** Reduce wordiness and ensure that the stem of the item sets the task as clearly as possible.
6. **Make Options Concise:** Reduce wordiness in options (often by relocating information into the body of the item).
7. **Reduce *If* Clause:** Convert complex sentences with an *if clause* and a *result clause* to two simple sentences.
8. **Simplify Verb Forms:** Change passive voice to active voice, and change past, future, or conditional tense to present tense.
9. **Reduce Wordiness:** Eliminate extraneous words or otherwise simplify and hone language used.
10. **Add Emphasis to Key Words:** Use underlining to draw students' attention to word(s) that highlights the task being set.
11. **Graphic Representation:** Revise or simplify the artwork or graphics associated with an item. This includes eliminating a graphic that does not help students understand the task in the item.

In order to make each item as linguistically accessible as possible, more than one category of modification was applied to most items. We tracked the modifications applied to each item so that analyses by categories could be conducted.

The development of these modification categories is more detailed than has been done for any other existing study. The study by Sato, Rabinowitz, Gallagher, and Huang (2010) is the closest in terms of the number of modification categories with five (context, graphics, vocabulary/wording, sentence structure, and format/style). Thus, one of the main contributions to the research literature on the linguistic modification of content assessments is the establishment of these categories as one possible approach to simplify language in assessments taken by ELLs.

## Analyses and Results

A total of 2,060 students, from more than 40 schools in 10 districts in two states, participated in the study (see Table 1). Data were analyzed separately for each content area and grade level.

### ***Effects of Linguistic Modification***

In order to assess the effects of linguistic modification, item means for proportion correct were computed by language status and test form. Thus, a 2×2 design was used to compare the performance of ELLs and non-ELLs on both the modified and original versions of each item. In addition, we used the total score on the six common NAEP items for each grade and subject as a covariate in an analysis of covariance (ANCOVA) design. The purpose of the covariate was to control for any pre-existing group differences across test forms. The ANCOVA results provided information on whether the interaction between the two independent variables—ELL status and test form—was significant. From the perspective of a group by treatment interaction, the best outcome would be one where the performance of the ELLs improved so as to be comparable to that of the non-ELLs, while the performance of the non-ELLs remained the same or improved slightly (an outcome sometimes referred to as *differential boost*).

The ANCOVA results showed that the effects of the linguistic modification for ELLs were generally mixed across grades and subjects. As shown in Tables 2 through 5, the *p*-values (percentage of students answering an item correctly) showed

**Table 3** Comparisons of *p*-Values by Language Status for Grade 4 Science

<i>p</i> -value comparison	English language learner	Non-English language learner
Original = Modified	9	9
Original > Modified	10	8
Original < Modified	11	13

**Table 4** Comparisons of *p*-Values by Language Status for Grade 6 Mathematics

<i>p</i> -value comparison	English language learner	Non-English language learner
Original = Modified	5	7
Original > Modified	11	9
Original < Modified	14	14

**Table 5** Comparisons of *p*-Values by Language Status for Grade 6 Science

<i>p</i> -value comparison	English language learner	Non-English language learner
Original = Modified	5	6
Original > Modified	11	12
Original < Modified	14	12

no consistent differences between the original and modified versions of the same item for ELLs. In these tables, if the *p*-values between the original and modified version differed by .02 or less (in order to take sampling error into account), we categorized the *p*-values as being equal. Out of the 120 items, ELLs performed better on the modified version of 48 items and did better on the original version of 41 items. In contrast, interestingly, the effects of linguistic modification appeared to be more positive for non-ELLs, as three of the four tables (all except Table 5) showed that there were more items in which the non-ELLs did better on the modified version than on the original version, rather than the other way around.

### **Categories of Linguistic Modification**

In order to answer the second research question, analysis by type of modification was conducted. For each content area, within each grade, items with the same type of modification were grouped together to form a category. Mean scores were computed and ANCOVA tests carried out to examine if there were significant differences in performance by category of modification. The one type of modification that showed the largest occurrence of significant results was #2 (Refine Context). It appears that ELLs benefited most from this modification, which led to a more accessible or simplified context for the test item.

### **Discussion**

This study added some mixed results to the literature on the effectiveness of the linguistic modification of test items for improving the assessment of content knowledge of ELLs. Using the expertise of a team of test developers and researchers, we systematically modified existing test items and confirmed with content specialists that the modified items faithfully assessed the construct of interest. Although the effects of linguistic modification were mixed, as is consistent with some earlier studies, we found that many modified items exhibited positive impact on the performance of ELLs, and, in particular, that the modification of Refining Context appeared to demonstrate positive benefits for ELLs on the mathematics and science assessments. For those items in which both groups of students performed better on the modified version, it may have been the case that the original version may have been unnecessarily complex. Because some non-ELLs may not be reading at grade level, these students may have also benefited from the simplified language used in the modified version of the items.

Among the possible explanations for our mixed findings is the possibility that the modifications we applied were simply not effective or that the students in the study were not in a position to benefit from the modifications (perhaps as a consequence of their English proficiency levels or some other background factor). The knowledge from this study may be useful as we seek to increase the validity of our content assessments for ELLs and for understanding the factors that contribute to the achievement gap for ELLs.

## Study 2: Cognitive Interviews of Students on Linguistically Modified Test Items

Because the first research study found mixed results from the effects of linguistic modification, a qualitative study using cognitive interviews was conducted as a follow-up to gain a greater understanding of how linguistic modification functions. These interviews were conducted with new samples of students from the target populations in order to shed light on the understanding and problem-solving strategies employed by students on the original and modified versions of the items. This study was designed to gain a better understanding of the strengths and limitations in applying linguistic modification to test items.

In theory, the linguistic modification of test items serves to clarify and simplify the language used in assessments, while preserving the construct being measured. For ELLs, it is an example of a testing accommodation that provides direct linguistic support for these students. However, as the first study reported here and other studies have shown, the results from modifying items are often mixed. The earlier study was conducted with more than 2,000 students, and we obtained sufficient quantitative information on the difficulty of the items for ELLs and non-ELLs. Because the quantitative analyses provided only limited information regarding the performance of students on these items, the cognitive interviews with students were intended to shed light on which features of the test items helped or hindered students in understanding and solving the items. A 50% sample of the items (or a total of 60 items) from the previous study that were easier, the same, or more difficult for students in the prior study was used in this study. We selected the items so as to be fully representative of the total set of 120 items with regard to group performance outcomes and modifications applied.

There are several different types of cognitive interviews currently used as data collection tools in the social sciences. In this study, we used a combination of concurrent reporting along with retrospective reporting in the cognitive interviews. Cognitive interviews based on direct probing (the use of retrospective questioning) are generally favored over the original (or pure) think-aloud method (solely concurrent thinking aloud), as they tend to show high reliability, generate data that are more suited to systematic coding and reporting, and provide a mechanism for obtaining information specific to the tasks presented (Beatty & Willis, 2007; Conrad & Blair, 2004). For a detailed comparison between the think-aloud and probing methods, see Beatty and Willis (2007) and DeMaio and Rothgeb (1996).

## Prior Research

In evaluating task comparability with respect to the content of a test and the test's items, studies that investigate the cognitive processes employed by examinees in responding to the items on a test can be particularly informative. The use of cognitive interviewing (sometimes referred to as *cognitive labs*) to obtain information on how students understand and solve test questions is increasing in popularity. Think-aloud methods have been used in psychological research to collect verbalization data on an individual's thinking and cognitive processes for more than 60 years (Duncker, 1945; Ericsson & Simon, 1993).

More recently, these methods have been used to evaluate students' performance and understanding on test items and to improve test design (Embretson & Gorin, 2001; Hamilton, Nussbaum, & Snow, 1997; Kopriva, 2001; Leighton, Rogers, & Mcguire, 1999; Martiniello, 2008; Thompson, Johnstone, & Thurlow, 2002). Kopriva (2000) and Winter, Kopriva, Chen, and Emick (2006) have outlined procedures for conducting cognitive labs with ELLs (Kopriva, 2008). Also with regard to ELLs, Young (2009) identified Test Content as one of the eight indicators of test comparability in his framework for test validity research on content assessments taken by ELLs. More specifically, studies are needed that investigate whether the content and cognitive processes used by examinees are the same across groups of students. Thus, the use of cognitive interviews is one mechanism by which evidence can be collected on the validity of test items used in content assessments for all students, including ELLs.

Johnstone, Bottsford-Miller, and Thompson (2006) conducted a study using think-aloud methods with students with disabilities and ELLs to determine whether fourth and eighth grade mathematics items had all of the elements of

universally designed assessments as specified by Thompson *et al.* (2002). Think-aloud methods do have their limitations: they may not work well for very difficult items, since highly proficient students work automatically, while less proficient students may not be able to verbalize why they do not understand the item. This limitation was first noted by Leighton (2004) and was evidenced in the Johnstone *et al.* (2006) study.

Martiniello (2008) used think-aloud protocols with 24 native Spanish-speaking ELLs to gauge their understanding of fourth grade mathematics word problems. Martiniello conducted think-aloud interviews with students individually in one or two sessions that lasted 30–90 min each. The interviews were conducted primarily in Spanish, although students who were more fluent in English used both languages. Students were shown the mathematics items one at a time, and Martiniello recorded students' decoding errors, as well as mistakes in interpreting complex text. Martiniello found that polysemous words (words with multiple meanings), specific home-based cultural references (such as coupon), and more sophisticated academic terms were especially challenging for ELLs. Martiniello also found that the ELLs were able to use their knowledge of Spanish-English cognates to decode the meanings of some words, but this seemed to be true only for high-frequency Spanish words that are likely to be familiar to children. Martiniello concluded that, "Witnessing the thinking process of making sense of math word problems can illuminate this process [of identifying item features that may cause DIF]" (Martiniello, 2008, p. 363).

## Research Methods

### Research Questions

For this study, the main research question is: which features of the linguistically modified items produced the observed outcomes for ELLs and non-ELLs? From the earlier study, we have information on the difficulties of the items for ELLs and non-ELLs. In this qualitative study, we wanted to identify and isolate the linguistic features of the items that produced the earlier results. Thus, by gathering data on the cognitive processes that students used to understand the items and by observing their problem-solving approaches, we hoped to identify the features that made the modified version of each item easier or harder to solve than the original version.

### Participants

The participants in this study were students in the fourth or sixth grade who were classified as either being an ELL or a native English speaker. Students were recruited from school districts in Delaware, Massachusetts, and New Jersey. Because the majority of ELLs are native Spanish speakers, we limited the ELL sample to only include native Spanish speaking students. In addition, the ELLs targeted for this study were at the intermediate levels of English proficiency (not basic or advanced), according to their state English language proficiency test score, since prior research studies have shown that linguistically modified items appear to have the most impact for students at these levels. It is worth noting that this is true when examining census and demographic data at the national level but is not necessarily true for each state. All three states that were part of the study are part of the World-Class Instructional Design and Assessment (WIDA) Consortium. The English language proficiency assessment in these states is the ACCESS for ELLs (Assessing Comprehension and Communication in English State-to-State for English Language Learners). This assessment has six levels: (a) Entering, (b) Beginning, (c) Developing, (d) Expanding, (e) Bridging, and (f) Reaching. We requested schools to recruit students who were in Levels 2–5. Despite our specific instructions, there were some cases where students who were recruited were not identified using the student's most recent ACCESS score, and interviewers were trained to dismiss students who did not know enough English to participate in the study.

The ELLs in the earlier study were also classified at the intermediate levels of proficiency. Lastly, it was hypothesized that students at the lowest level of English proficiency would not be able to interpret enough of the test items that were presented in English to produce useful data and, therefore, we determined not to be appropriate study participants.

The original target sample size for this study was 180 students (45 for each language classification and grade). After the completion of several interviews, there were 11 cases that were not included in the final sample, due to the low English proficiency of the participants. The final sample size was 159 students, which was 88.3% of the targeted sample. The inability to reach our expected sample size was due to a difficulty in recruiting and the removal of the 11 unqualified students. See Table 6 for the breakdown by language classification and grade of the number of student participants.

**Table 6** Student Participants by Language Classification and Grade

Language classification	Grade 4	Grade 6
English language learner	42	33
Non-English language learner	41	43

### ***Instruments***

We used a total of 60 items, in two content areas (mathematics and science) and two grade levels (third and fifth). Note that because students were interviewed at different times during the school year, we decided that it was best to use items from a grade lower than the students' current enrollment to avoid the problem of students not having learned the subject matter previously. Thus, fourth grade students were shown items from third grade tests, while sixth grade students were shown items from fifth grade tests. Each of the items in this study was represented in an original and modified version. The items were bundled into test booklets with four item pairs (original and modified version of each item) per booklet; each student received one booklet to reference during the interview.

### ***Interview Protocol***

Interviewers followed a protocol to guide each cognitive interview session. See Appendix A for a partial example of the interview protocol that was developed and used in this study. The protocol includes the same set of questions for each pair of items; only questions for the first pair of items are shown. There was one protocol printed for each interview, which was accompanied by a corresponding test booklet for the student and separate sheets of the items to use for the comparison questions. Students recorded their answers to the test items in their respective test booklets and responded to the retrospective and comparison questions orally. The protocol presumed the students were already in the testing room and began with seating instructions. Information about the test and purpose of the study were described, and the digital voice recorder was checked for effectiveness. The interviewer would demonstrate thinking aloud with a math or science question that was similar in appearance and difficulty level as the items in the test booklet, and the student would get a chance to practice thinking aloud with another item similar to those in the test. The research session of the interview followed after the practice. There were three components that guided the research session, which are described below. Specific information to say to the student was provided, but interviewers were encouraged to speak conversationally. Instructions and hints to the interviewer were also described in the protocol.

#### *Concurrent Notes*

As the student thought aloud, the interviewer had an image of the same item the student was solving on the left page of the protocol and a concurrent notes page divided into categories on the right page. Interviewers were instructed to take notes on the item content, item format, and answering strategy while the student was coming to his or her answer. They were to remain silent as the student solved the item, unless the student was not speaking. If needed, the interviewer would gently remind the student to think aloud.

Each item was provided in the protocol as the student saw it to allow the interviewer to make a quick note on any aspect of the item (e.g., the stem, part of an image, one or more words) that a student had an issue or mentioned specifically. Interviewers were asked to include information about any errors in the content (math or science) and any issues understanding the question, stem, or specific words. Information about the format, such as a student reference to pictures, or the layout of the item or information on the page was to be noted. Interviewers were also asked to document during or immediately following the session whether the student was guessing or was solving each problem using appropriate test-taking strategies. They were to summarize the steps the student followed and document whether the strategy was the same for each of the items in the pair. There was also an open notes area for any other noteworthy comments related to answering the item or about the session or the student, such as a noticeable temperament (e.g., distraction, engagement, frustration).

#### *Retrospective Questions*

As soon as the student finished answering an item, the interviewer would ask five questions to further probe for understanding. These questions were asked after each of the eight items was delivered to the students. Below are the five retrospective questions and the purpose for asking each.

1. *Do you think you got this question correct and why?* This question was asked to get a sense of confidence in a student's answer and to determine whether the student knew he or she was correct, regardless of whether this was an accurate estimation.
2. *Was this question hard, in the middle, or easy for you and why?* This was asked to gauge the student's perception of the level of difficulty for the item and to see whether the students would vary in their perceptions among the pairs of items.
3. *Is this something you have learned in your mathematics (science) class before?* This question was asked to determine whether the student had the opportunity to learn the information. If this question was interpreted correctly, the goal of the question was to capture if the student had learned the content, regardless of whether it was in the United States or in his or her native country.
4. *Were there any words, ideas, or anything else that made it easy to answer this question?* This question was asked to identify any parts of the question that the student found to be helpful to answer the item. The intention was to think about all aspects of the item, such as the specific context the item used, the specific wording, or the layout of the item.
5. *Were there any words, ideas, or anything else that made this question harder or confusing to answer?* This question was asked to complement the prior question to identify components of the item that made it more difficult or did not allow the student to answer the item.

### Comparison Questions

As soon as the student finished answering the second item of each pair of items, the student was asked three comparison questions. The interviewer would show the student two sheets displaying the pair of items that were just answered. The three comparison questions and the rationale for asking each are described below.

1. *Which question lets you know what you are supposed to do and why?* This was asked to get a sense of whether one of the items in the pair was more clear or understandable. The question was intentionally open-ended so as not to influence the particular aspects of the item that made it clearer or to guide the level of detail that the student decided was appropriate to provide to describe his or her choice.
2. *Which of the two questions do you like better and why?* This question was asked to determine if the student preferred one item of the pair. This question was also left open-ended so the student could identify any reason for a preference, including the format, wording, particular scenario, or a combination of reasons. This was asked in addition to the first comparison question because it was hypothesized that students would not always select the same item of the pair when asked these two questions separately.
3. *Do you think that these two questions are asking you about the same mathematics (science)?* If yes, then ask for both items at once; if no, then ask for each item version specifically: *If I were a student in your class and I did not understand what this question was asking, could you explain what this question is asking me to do in your own words?*

The third question was asked to collect an overall measure of the student's understanding of the meaning of each question. A complete response to this question would provide information to demonstrate whether the student knew the content and what the question was asking or would identify areas where the student had trouble. This was intended to be supplemental to the think-aloud response. It could reiterate what was stated as the student spoke his or her thoughts and final answer choice, or it could provide additional information about the student's understanding or lack of understanding of the item that was not uttered during the initial think-aloud.

### Interview Procedures

Interviewers met with students one-on-one during the school day to conduct a cognitive interview that lasted about 1 hr. Students were shown up to four pairs of items in each interview (each pair of items consists of the original and modified versions). Item versions were presented using a counterbalanced design so that the order of presentation of the two versions (original or modified) of a particular pair of items alternated from one test booklet to the next; this design was used because it controls for possible order effects. There were 13 interviewers who completed a half-day training session to learn how to conduct the interviews and how to enter data after each session was completed. Each interview with the native English speaking students was conducted in English. All of the interviewers of ELLs were bilingual English/Spanish

speakers. The interviewers who met with the ELLs instructed the students to speak in whatever language they were thinking. Most information could be provided to the students in Spanish if the participant asked a clarification question, with the exception of anything that affected the construct if it were translated to Spanish. The item stem and answer options were not to be translated into Spanish. Each interview was audio recorded and later transcribed. The ELL interviews were directly transcribed, and a translated-to-English version of the transcription was also produced to allow native English speaking coders to understand the interviews. Interviewers also recorded their notes from each session into an Excel database.

### **Data Analysis**

The qualitative data collected in this study consisted of interview notes taken during the cognitive interviews. The data were entered, coded, and analyzed using the NVivo 8 software for qualitative analysis. The software was used to organize the data, provide summaries, explore trends, and build and test theories regarding students' responses to the test items. By linking students' responses to specific items and item features, we were able to gain a better understanding of how the linguistically modified items produced the observed outcomes for ELLs and non-ELLs.

We also summarized students' responses using Excel spreadsheets as an approach to organizing the data in a systematic way (see Appendix B for an example). For each grade and subject, we created spreadsheets organized by student identification codes and listed the relevant features for each response including:

- The type of modification for a particular item,
- Whether the student answered the item correctly or not,
- The strategies the student used in answering the item,
- Whether the student believed he or she answered correctly,
- Whether the student had already learned the information tested in the item,
- Whether the student preferred one version of the item or not, and
- Whether the student believed that he or she could explain the item to a classmate.

In addition, we created tables to summarize the information contained in the Excel spreadsheets; in Appendix C, we have included four summaries, one for each grade and subject. Given the volume of data, summarizing the information in a systematic manner can be extremely helpful in understanding students' responses and in identifying significant patterns and trends.

### **Results**

As is evident, a substantial amount of data was collected for this study. In the appendices to this report, we have included selected samples of the original data and summaries we produced to facilitate our analyses. Readers interested in inspecting the entirety of our data can do so by contacting one of the authors of this report. In order to provide some structure to the results from this study, we have organized our findings according to the analyses we conducted to answer the main research question.

We conducted our initial analyses of students' responses using the student interview transcripts along with the relevant data coded into NVivo. Our analyses were carried out within grade by subject area and then by specific linguistic modification within each grade by subject area. In general, we found that there were minimal differences across type of linguistic modification with regard to the overall effects on students' understanding or performance. This finding was true for both the ELLs and the non-ELLs in this study. Further, we found that modifying an item generally yielded improvements in the desired direction for ELLs. That is, ELLs more often reported that the modified version of an item was easier to understand, had fewer difficult words or concepts, and was preferable when compared with the original version of the same item.

However, these effects were not uniformly consistent across the modifications and items used for this study. Although the desirable effects resulting from the use of modified items on ELLs' understanding and performance were in a positive direction, these effects were neither strong enough nor consistent enough to warrant any definitive conclusions. Students' responses during the cognitive interviews indicated that the modified version of an item generally had more of the positive attributes that enhanced students' understanding. However, students also reported (although not as frequently as with the modified version of an item) the same qualities for some of the original versions of the items. When differences between the two versions of an item were identified, the modified version of the items produced the desirable outcomes, on average,

The chart below shows the number of each kind of kite that are mixed up in a box.  
Kites for Sale

Kind of Kite	Number of Kites
	4
	7
	18
	11

If Cathy picks a kite from the box without looking, what kind is it MOST LIKELY to be?

- A 
- B 
- C 
- D 

**Figure 1** A mathematics item administered to fourth graders. Example of modified version helped understanding of item: simplifying vocabulary.

about 60% or more of the total occurrences. It should be noted, however, that there were many instances where there were essentially no differences in students' understanding or performance on either version of an item. Thus, the figure given is actually an overestimate of the impact of linguistic modification, as there were many items for which the effects of modification were negligible or mixed. In addition, the mixed findings across modifications and items were as likely to be true for the native English speaking participants as for the ELL students.

### **Modified Version Helped Understanding of Item: Simplifying Vocabulary**

For example, a mathematics item given to fourth graders used different objects in the stem of the question. In the original version, the word *kites* was used, while in the modified version, the word *balls* was substituted (see Figure 1). An ELL student did not appear to know what kites were; when asked which version of the item he preferred, he chose the modified one since "... this one is like easier to say and this one [the original version], um a little bit hard." For each version of this item, the student chose a different incorrect answer option. In addition, he reported that he thought the two versions of the items did not ask about the same mathematical concept. He also reported that he did not think that he could explain either version of the item to another student in his class. In this case, the modification used of simplifying vocabulary, by replacing an unfamiliar word with a more accessible one, made the item easier for the student to read but did not lead him to answer the item correctly.

### **Modified Version Helped Access to Math Concepts: Making Stem Concise**

In a similar example (see Figure 2), a Grade 6 mathematics test item tested for approximating volume conversions, the original item asked the question, "Tina used 1 liter of chicken broth to make her special rice dish. This amount of chicken

**Tina used 1 liter of chicken broth to make her special rice dish. This amount of chicken broth is *closest to*—**

- A 1 ounce
- B 1 cup
- C 1 pint
- D 1 quart

**Figure 2** A mathematics item administered to sixth graders. Example of modified version helped access to math concepts: making stem concise.

broth is closest to ...” The modified item removed the empty context and made the stem concise by asking, “One liter is closest to ...” One ELL student was presented the original item first and did not know what the term *chicken broth* meant, an instance that supports the Martiniello (2008) finding that words used in the home are more likely to be known in a student’s native language and not necessarily in English. He understood enough about the concept of a liter and the answer options to pick an answer. He selected the same answer for the modified item and recognized that both questions asked the same question. In this case, despite the fact that the student did not answer either item correctly, it was encouraging to see that the more concise stem eliminated the problem of not understanding all of the words in the item.

### ***Modified Version Helped Understanding of Item: Remove Empty Context***

As another example, a different mathematics item (see Figure 3) that was given to fourth graders used different objects in the stem of the question. In the original version, the words *string* and *beads* were used. In the modified version, these words were eliminated, as the modification applied was to remove empty context. An ELL student was confused by these words in the original version, but she found the modified version much easier to understand. She reported that she thought the two versions of the items asked about the same mathematical concept, and she answered both versions correctly.

### ***Contradictory Finding: Making Stem Concise***

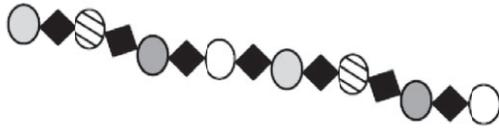
One example of a sixth grader answering a science item (see Figure 4) is as follows. The stem of the question asked which chemical element is present in all living things. In the original version of this item, the student selected *sodium*, which is incorrect; however, in the modified version, which employed the modification of making the item stem concise, the student selected the correct response, *carbon*. Yet, despite getting the item correct in the modified version, the student reported that the original version was both preferable and easier. This is one example of the type of seemingly contradictory results that was found during the course of our analyses.

### ***Contradictory Finding: Refining the Context of Item***

In an example of a fourth grader answering a science item (see Figure 5), the outcomes that resulted were again mixed with regard to the effects of applying linguistic modification. The stem of the question asked what caused a balloon that is left by a window for an hour to become larger. In the original version of this item, the student selected *heat from the sun*, which is the correct option. However, in the modified version, which employed the modification of refining the context of the item, the student selected an incorrect response, *the string attached*. Yet, despite getting the item correct in the original version, the student reported that the modified version was both preferable and clearer. Again, this is another example of the type of seemingly contradictory results found during the course of our analyses.

As was shown in these few examples, the information collected through the cognitive interviews was often contradictory to the outcomes observed, so that additional clarity and greater understanding regarding the effectiveness of the

**Look at these beads. The pattern repeats the first eight beads over and over.**



**The pattern continues in the same way. What are the next two beads?**

- A
- B
- C
- D

**Figure 3** A mathematics item administered to fourth graders. Example of modified version helped understanding of item: remove empty context.

**All living things contain which element?**

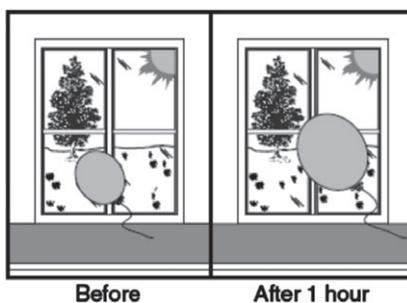
- A Helium
- B Sodium
- C Copper
- D Carbon

**Figure 4** A science item administered to sixth graders. Example of contradictory finding: making stem concise.

different linguistic modifications was limited. The data we collected for this study were analyzed systematically and exhaustively, but, unfortunately, the information we obtained through the cognitive interviews did not markedly increase our understanding of the findings from the previous research study. We did not develop any better understanding, beyond what was learned in the prior study, of which types of modification, if any, are most effective in increasing the comprehension or performance of ELL students on these content items.

**Discussion**

The primary goal of this research study was to determine, through cognitive interviews with students, which linguistic modifications can produce improved performance and understanding for ELLs, primarily, and for non-ELLs, secondarily. Unfortunately, the findings from these cognitive interviews added little clarity with regard to which of the linguistic modifications were effective in improving the understanding and performance of ELLs on mathematics and science items. The interview data that were collected and analyzed varied in the quality of valid information that could be extracted from the students’ responses. In addition, the student participants were not consistent, either within themselves or across students, in the information that they provided. Thus, the interviews conducted provided little additional insight, beyond



**Bobby put his balloon near a window. The picture shows what happened to the balloon after it was near a window for about an hour. Which of these probably caused the balloon to become larger?**

- A** Heat from the sun
- B** Glass in the window
- C** The tree by the window
- D** The string on the balloon

**Figure 5** A science item administered to fourth graders. Example of contradictory finding: refining the context of item.

what was already learned in the original study, regarding which of the linguistic modifications were most effective and the mechanisms by which they increased the understanding and performance of ELLs and non-ELLs on these content items.

These mixed results are consistent with findings from other studies of the use of linguistic modification to improve accessibility on content items for ELLs. In theory, the simplification of language should be an effective testing accommodation for ELLs; however, the research studies conducted to date provide limited empirical support for this approach. It is possible that because we modified existing test items rather than creating linguistically simplified items from scratch, any modification applied would be limited in its effectiveness. This approach was chosen because it more accurately reflects the typical test development process for modifying items. In addition, we needed to have both original and modified items in order to carry out comparisons between the two versions. Thus, retrofitting already existing items was the best approach for achieving these goals.

Although the results from this study are disappointing with regard to the systematic application of linguistic modification as a testing accommodation for ELLs, the experiences gained from these studies, particularly in the development of rubrics and processes for the various categories of modification, are valuable as future research tools and as recommended best practices for developing content assessments that are of greater accessibility to ELL students.

### Lessons Learned

We have included this section in our article as we learned quite a bit about conducting cognitive interviews with ELL students during the course of this study, and we thought it would be important to capture this information for the benefit of other researchers.

### Recruiting Schools

Finding schools to participate was a pervasive problem throughout this research study. Despite offering a financial incentive to the school for each student who participated, we had very few schools among those invited that were willing to allow the study to be conducted. We attempted to be flexible and offered to conduct the interviews during or after the

school day. It was preferred to conduct the interviews during the school day at the schools that agreed to be part of the study. One possible way to increase the number of interested schools could have been to offer significantly larger incentive to the schools. The expected goal would have been to reduce the amount of administration time spent on the project trying to get schools to participate. It may have also been helpful to initially obtain the approval to conduct research at the state or district level before inviting individual schools. Because we had a specific group of students that we were interested in recruiting, another possible solution would have been to build or use existing relationships with Hispanic professional organizations and interest groups to help promote our research study in schools by providing letters of support.

### **Conducting Interviews With English Language Learners**

It became apparent that some of the native Spanish speaking students that were recruited were not appropriate to participate in this study. Our initial criteria had the goal of collecting data from students with a range of English proficiency. We recruited students who were at the second to fifth (or intermediate) levels of proficiency. For the WIDA English proficiency test, the first level of proficiency is entering, the second level is beginning, the third level is developing, the fourth level is expanding, and the fifth level of English proficiency is bridging. In some cases, it was clear that the ELL students did not understand enough English to meaningfully interpret the test. This was also reflected in the notes that the interviewers wrote, which ranged from students not understanding specific words to not understanding the whole question. In future studies, it would be prudent to show the level of the English in the test to teachers to help identify any ELL student who is on the low end of the proficiency spectrum that would not be able to understand the English text. In addition, showing the content to the teachers to get a sense of whether their students have received an opportunity to learn the material would be helpful, even though it was asked in the interview if the student had previously learned the material. Unless the purpose of a study permits a test to be translated to a student's primary language, we did not find the procedure of cognitive interviews to be appropriate for ELLs at the beginning level. It is still important to capture feedback from students with the lowest level of English proficiency, since excluding part of the intended sample makes any research findings less generalizable. One alternative to a structured cognitive interview is to gather feedback from a basic conversation with the student in his or her native language about the items, the student's understanding of the English text in the items, and the student's opportunity to learn the content in his or her native country or in the United States. The best ways to collect information about items and accommodations from the students at the lowest levels of English proficiency warrants further investigation.

### **Using Fewer Item Pairs in the Cognitive Interviews**

In retrospect, this study would have been conducted more effectively if fewer item pairs were used in the study. By doing so, the sample size would have been reduced, and more focused attention could have been given to increase the quality of the data collected. Because some of the reviewers were concerned about the generalizability of the findings (which is not a primary goal of most studies employing cognitive interviews), an overly large number of item pairs were required for the study. We originally proposed to use 30 item pairs; the reviewers required an increase in the number of item pairs to 60. However, these particular reviewers were unfamiliar with the main purposes of cognitive interviews and imposed an overly stringent set of requirements on the study's research design.

### **Training Techniques for Interviewers**

Initial training of interviewers included a review of the purpose of the study; a detailed description of the materials and how to use them; a review of the postinterview steps, including the database to enter postinterview notes; and time to practice administering the protocol. It was evident after examining the first round of interview transcripts that some of the interviews were incorrectly administered, so these data were removed from our analyses. After a review of the errors that occurred during the first round of interviews, additional training was given to all interviewers, which included describing appropriate follow-up questions to the retrospective and comparison questions and note taking techniques to ensure the data summary files were sufficiently thorough. One lesson was to instruct interviewers to remind students to speak aloud everything, rather than nod in agreement or disagreement or point to one specific location in a test booklet because that information would not be captured on the audio recordings. It was found to be helpful to have longer training sessions

that included how to appropriately take notes during an interview, a description of the level of specificity to include in the summary notes, and time to practice the interview as a group demonstrating appropriate follow-up questions that may not be specified in the protocol.

### Modifying Protocol Questions

Despite having a specific purpose for each of the retrospective and comparison questions, the responses from students varied noticeably. This could have been due to multiple interpretations of the question or the students not understanding the question. After the analysis of the interviews was conducted, we learned that some questions produced redundant information, and others elicited responses from students that may or may not be accurate despite pilot testing the protocol questions. The question “Is this something you learned in your mathematics class before?” was interpreted literally by some students as to mean the same mathematics item within the same context of the item in the test was taught in his or her class. The students would say they did not learn it, but would then explain that they were taught the subject by their teacher using another example. It was important for the interviewer to appropriately explain this misinterpretation of the question in the summary notes. Having a system for interviewers to relay when interview questions are misinterpreted or not working is important. It is also recommended that future cognitive interview training includes sufficient explanation of the purpose of each question to allow interviewers to restate questions that the students may not understand as they are worded in the protocol. This is especially important for interviews with ELL students when the interviewer is permitted to translate questions in the student’s primary language.

### Presenting Two Versions of an Item to the Same Student

Prior to this study being conducted, there was not an example of a study that provided an appropriate method for collecting individual feedback on multiple versions of the same item. The researchers asked colleagues in the field for their input on the study design. It was determined that it was best to present both versions of the items to the same student so that we could gather comparative information. One other important lesson learned from the administration of the test booklet questions was the fact that a number of students recognized that the item pairs were asking the same question, before being asked if the questions were asking about the same topic. Even though the order of the questions presented varied, a more random presentation could alleviate these presentation order problems in the future.

### Conclusion

In conclusion, these two research studies found mixed results for the impact of linguistic modification of test items from K–12 mathematics and science content assessments. Although the modifications were developed based on the expertise of a team of researchers and assessment developers, the results of the quantitative study for ELLs were inconclusive. The follow-up study using cognitive interviews of students led to some additional insights regarding the effects of linguistic modification, although these findings were not as clear as we had hoped. Nevertheless, we believe that these two studies have provided useful information about this testing accommodation for ELLs, and we encourage other researchers to continue investigating and improving the efficacy of linguistic modification of test items on content assessments taken by ELLs.

### References

- Abedi, J. (2007). *Language factors in the assessment of English language learners: The theory and practice underlying the linguistic modification approach*. Manuscript prepared for the U.S. Department of Education, LEP Partnership.
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14, 219–234.
- Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners’ test performance. *Educational Measurement: Issues and Practice*, 19, 16–26.
- Aguirre-Muñoz, Z. (2000, April). *What does knowledge look like? Mapping English language learners’ conceptual knowledge*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71, 287–311.

- Burstein, J., Shore, J., Sabatini, J., Moulder, B., Holtzman, S., & Pedersen, T. (2012). *The Language Muse system: Linguistically focused instructional author* (Report No. RR-12-21). Princeton, NJ: Educational Testing Service.
- Conrad, F., & Blair, J. (2004). Data quality in cognitive interviews: The case for verbal reports. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lesser, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 67–87). Hoboken, NJ: John Wiley & Sons.
- DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In S. Norbert & S. Seymour (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 177–195). San Francisco, CA: Jossey-Bass.
- Duncker, K. (1945). On problem-solving. In J. F. Dashiell (Ed.), *Psychological monographs* (pp. 1–114). Washington, DC: American Psychological Association.
- Embretson, S., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38, 343–368.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). Cambridge, MA: MIT Press.
- Hamilton, L. S., Nussbaum, E. M., & Snow, R. E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, 10, 181–200.
- Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). *Using the think aloud method (cognitive labs) to evaluate test designs for student with disabilities and English language learners* (Technical Report No. 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79, 1168–1201.
- Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000–2001 summary report*. Washington, DC: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.
- Kopriva, R. J. (2000). *Ensuring accuracy in testing for English language learners: A practical guide for assessment development*. Washington, DC: Council of Chief State School Officers.
- Kopriva, R. J. (2001, June). *ELL validity research designs for state academic assessments: An outline of five research designs evaluating the validity of large-scale assessments for English language learners and other test takers*. Paper presented at the Council of Chief State School Officers Meeting, Houston, TX.
- Kopriva, R. J. (2008). *Improving testing for English language learners*. New York, NY: Routledge.
- Lee, J., Grigg, W., & Dion, G. (2007). *The nation's report card: Mathematics 2007* (NCES 2007-494). Washington, DC: National Center for Education Statistics.
- Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007* (NCES 2007-496). Washington, DC: National Center for Education Statistics.
- Leighton, J. P. (2004). Avoiding misconceptions, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, 23, 6–15.
- Leighton, J. P., Rogers, W. T., & Mcguire, T. O. (1999). Assessment of student problem solving on ill-defined tasks. *Alberta Journal of Educational Research*, 45, 409–427.
- Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review*, 78, 333–368.
- National Clearinghouse for English Language Acquisition. (2006). *The growing numbers of limited English proficient students: 1994/95–2004/05* [Poster]. Washington DC: Author.
- No Child Left Behind Act. (2002). No Child Left Behind Act of 2001, 20 U.S.C. § 6301 et seq.
- Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice*, 30, 10–28.
- Sato, E., Rabinowitz, S., Gallagher, C., & Huang, C.-W. (2010). *Accommodations for English language learner students: The effectiveness of linguistic modification of math test item sets* (NCEE 2009-4079). Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Center for Educational Assessment Research Report No. 485). Amherst, MA: University of Massachusetts.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report No. 44). Minneapolis, MN: National Center on Educational Outcomes.
- Winter, P., Kopriva, R. J., Chen, S., & Emick, J. (2006). Exploring individual and item factors that affect assessment validity for diverse learners: Results from a large-scale cognitive lab. *Learning and Individual Differences*, 16, 267–276.
- Young, J. W. (2009). A framework for test validity research on content assessments taken by English language learners. *Educational Assessment*, 14, 122–138.

## Appendix A

### Linguistic Modification Cognitive Interview Protocol

#### COGNITIVE LAB INTERVIEW PROTOCOL

##### SEATING INSTRUCTIONS



Follow the school's procedures to escort one student from his or her classroom to an empty testing room or corner of a quiet room.

Make sure that you and the student are seated so that you face one another directly across the table or desk if possible. Set the interview protocol and the stopwatch in front of you and the closed student test booklet in front of the student. Place the digital recorder on the table near the student.

##### INTRODUCTION

Hello! My name is (YOUR NAME). I work at Educational Testing Service. Today you will be participating in a special study on math. Its goal is to find better ways to test math for all students. We need your help to pick the best test questions. If at any point you decide you do not want to continue that is your choice and you are free to stop and go back to class.

Do you have any questions before I begin the instructions?



Create small talk...ask a couple friendly questions such as the following:

Do you like math?



If yes, say

Oh, good, then I think you will like what we are doing today.



If no, say

Okay, well what is your favorite subject?



Appropriately respond. (Also make a note in the test administrator answer sheet whether student likes to read.)

Then move go back to the directions. Say,

You will not be asked to write your name on any of the work you do, and *no one* in the school will see your answers. The information you provide will be combined with information from other students from New Jersey, Pennsylvania and Massachusetts. Because the study is so important to schools and students, I want to thank you ahead of time for all the hard work you are about to put into your answers. Some of the questions might be difficult, but just do the best that you can.

#### AUDIO RECORDING/SOUND CHECK

We are going to record today's test so I remember what you say. First we're going to test out this audio recorder and the microphone.



Make a test recording by telling the student:

When I say GO, I want you to say, "Today is Friday" in a clear voice.



Press the RECORD button on the audio recorder and say "YOU MAY BEGIN" to prompt the student.

Be sure to press RECORD *before* you tell the student to begin so your voice is recorded too.

If necessary, remind the student to say the sentence ("Today is Friday.") Once the student has said the sentence, press PLAY to listen to the recording.

- If you *cannot* hear the student clearly on the recorder, make sure the microphone is plugged into the proper jack on the recorder and repeat the test recording process until you get a clear recording, then continue with the administration:
- If you *can* hear the student clearly, continue with the administration:

Good job. Now we'll begin the instructions.

## DIRECTIONS



### Think aloud hints to administrator

We're interested in capturing all the mental processes that the student engages while completing these math items. To do this your goal is to have the student speak aloud all his or her thoughts while solving the items and by asking follow up questions for each item. There are several things to keep in mind to ensure the data collected is as complete as possible:

- If a student is continually providing short responses or not answering, use “continuers” to encourage the student to be more descriptive. The trick is to get them to continue without putting words in their mouth. Make sure not to ask questions that lead a student's response. You have to be as objective and non-bias as possible. The way you get them to keep talking is by offering a gentle verbal nudge like this:

“What are you thinking now?”

“Any other thoughts?”

“Tell me how you came to pick that answer.”

- Use your best judgment. If a student is responsive but is not easily explaining his or her reasoning, gently probe the student without putting any bias in their response.
- Focus on the task at hand, the particular item. Do not try to ask a student a question in general terms.
- It's important to capture the frustration aspect of taking the test (and performing the think aloud exercise) if it seems a student is hung up with something.

## INSTRUCTIONS

## DIRECTIONS

You are going to be answering 8 math test questions. Please open your test booklet to the first page where you will see the directions. To mark an answer, circle the letter next to your choice.



Direct student to the first page and point to the correctly circled answer.

**PRACTICE EXAMPLE AND DETAILS**

Today I will show you some mathematics test questions and ask you to talk to me as you answer them. I want you to read the question aloud and then tell me what you are thinking about while you answer it. In a moment, I will show you how to think out loud as if I were a student, and then I will give you a chance to practice thinking out loud before we begin the test.

We are doing this activity so you can help us make the test questions better for students in your grade. You are not being tested; we are trying to make the test questions better. So there are no wrong answers and anything you say could help us.

Please turn the page in your booklet so you are at the first practice page.



Make sure student is on the administrator practice page.

I'm going to show you how to answer a question using the think-aloud process. You can follow along in your book to see the question I am answering. When I am finished, you will get a chance to practice.

**ONLY FOR ELLS** – When you are talking out loud, please say your thoughts and answer in English or Spanish, whatever is easier for you to say all your thoughts out loud.



Make sure student is on the correct page.

Read item and answer choices aloud and demonstrate think aloud

**Mike went into the library at 10:12 a.m. He left the library at 5:43 p.m. How long was he in the library?**

- A 7 hr 31 min
- B 7 hr 55 min
- C 15 hr 31 min
- D 15 hr 55 min

- As you heard, I was talking the whole time I was working on the question.
- I began by reading the question and the answer choices aloud, and then
- I said aloud everything I was thinking about as I answered the question.
- I also made a mistake and had to start over, which is okay because sometimes the answer doesn't come quickly to me and I have to try several things before I can answer the question.

Do you have any questions?



Answer any questions the student may ask

Now I'd like you to practice answering a question by thinking aloud. Please read the question and the answer choices aloud first, and then I'd like you to say all of your thoughts aloud as you answer the question.



Direct student to the student practice question

Student practice think aloud

**Tracy walked 2.31 miles. Lisa walked 1.95 miles. How much farther did Tracy walk than Lisa?**

- A 1.66 miles
- B 1.36 miles
- C 0.46 miles
- D 0.36 miles

[According to the student's performance, assist with appropriate suggestions for more accurate thinking aloud or say] Good job! Do you have any questions?



Answer any questions the student may ask.

Okay, let's move on to the actual test questions. I want you to read each question one and say all your thoughts out loud as you answer the question. Then I will ask you some questions about it. Don't move onto the next question until I tell you. We will do this together for a total of 8 questions.

One thing that you can do that will help me is to circle any word that you read that you do not understand or think is confusing.

So that I can remember what you say, I am going to be taking notes while you think aloud and answer the questions. Remember, you are not receiving a grade on the questions you answer today.

**REMINDER:** Press the RECORD button to begin recording and continue with the administration.

[SAY] The student ID and the date.

Please turn the page in your booklet to the first test question and you may begin reading aloud and answering the question.

1

Irene has a group of counters like the ones pictured below. In the group are 9 counters with a star on them and 1 counter without a star.



If one counter is chosen from the group without looking, which of the following best describes the chances that it will be the one *without* a star?

- A Certain
- B Likely, but not certain
- C Unlikely, but not impossible
- D Impossible

**Concurrent reporting notes page**

#1

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

**Other noteworthy comments:**

1

Irene has a group of counters like the ones pictured below. In the group are 9 counters with a star on them and 1 counter without a star.



If one counter is chosen from the group without looking, which of the following best describes the chances that it will be the one *without* a star?

- A Certain
- B Likely, but not certain
- C Unlikely, but not impossible
- D Impossible



5. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (indicate the words if Yes)

Great, thank you, now I'd like you to answer another question.

2

**Irene has 10 cards. She has 9 red cards and 1 blue card. Irene will pick one card without looking at it. What are the chances that she will pick a blue card?**

- A Certain
- B Likely, but not certain
- C Unlikely, but not impossible
- D Impossible

**Concurrent reporting notes page**

#2

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

**Other noteworthy comments:**

2

**Irene has 10 cards. She has 9 red cards and 1 blue card. Irene will pick one card without looking at it. What are the chances that she will pick a blue card?**

- A Certain
- B Likely, but not certain
- C Unlikely, but not impossible
- D Impossible

Now I'm going to ask you some questions about the question you just answered.

### Retrospective Questions

1. Do you think you got this question correct? Yes or No
  - a. Why or why not?
  
2. Was this question hard, in the middle or easy for you?
  - a. Why?
  
3. Is this something you have learned in your math class before?  
Yes or No or I don't know (*or indicate other response*)
  
  
  
  
  
  
  
  
  
  
4. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
  
  
  
  
  
  
  
  
  
5. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

Now I'm going to ask you some questions about the two questions you just answered that were similar.

### Item Comparison Questions

1. Which question lets you know what you are supposed to do? First, second, or no preference?
  - a. Why?
  
2. Which of the two questions do you like better? First, second, or no preference?
  - a. Why?
  
3. Do you think that these two questions are asking you about the same math? Yes or No
  - a. If Yes, then ask for both items at once,
    - i. If No, then ask for each item version specifically:
      - ii. If I were a student in your class and I did not understand what this question was asking, could you explain what the first question is asking me to do in your own words?
      - iii. If I were a student in your class and I did not understand what this question was asking, could you explain what the second question is asking me to do in your own words?

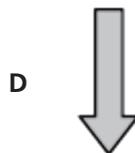
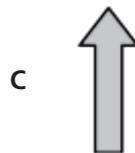
Okay, let's move onto the next question.

3

The first ten shapes in a pattern are shown below. Lionel is making the pattern by repeating the first six shapes in the same order as shown in the picture.



If Lionel continues the pattern in the same way, what will be the 16th shape in the pattern?



**Concurrent reporting notes page**

#3

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

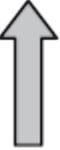
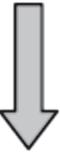
**Other noteworthy comments:**

3

The first ten shapes in a pattern are shown below. Lionel is making the pattern by repeating the first six shapes in the same order as shown in the picture.



If Lionel continues the pattern in the same way, what will be the 16th shape in the pattern?

- A 
- B 
- C 
- D 

Now I'm going to ask you some questions about the question you just answered.

### Retrospective Questions

1. Do you think you got this question correct? Yes or No
  - a. Why or why not?
  
2. Was this question hard, in the middle or easy for you?
  - a. Why?
  
3. Is this something you have learned in your math class before?  
Yes or No or I don't know (*or indicate other response*)
  
4. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
5. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

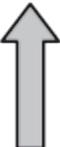
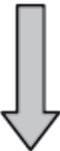
Great, thank you, now I'd like you to answer another question.

4

Look at the pattern. The pattern repeats the first six shapes in the same order.



What will be the 16th shape?

- A 
- B 
- C 
- D 

**Concurrent reporting notes page**

#4

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

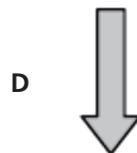
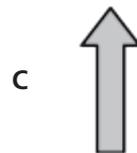
**Other noteworthy comments:**

4

Look at the pattern. The pattern repeats the first six shapes in the same order.



What will be the 16th shape?





Now I'm going to ask you some questions about the two questions you just answered that were similar.

### Item Comparison Questions

1. Which question lets you know what you are supposed to do? First, second, or no preference?
  - a. Why?
  
2. Which of the two questions do you like better? First, second, or no preference?
  - a. Why?
  
3. Do you think that these two questions are asking you about the same math? Yes or No
  - a. If Yes, then ask for both items at once,
    - i. If No, then ask for each item version specifically:
    - ii. If I were a student in your class and I did not understand what this question was asking, could you explain what the first question is asking me to do in your own words?
    - iii. If I were a student in your class and I did not understand what this question was asking, could you explain what the second question is asking me to do in your own words?

Okay, let's move onto the next question.

5

Each person running in a race will get a card with a number on it. Look at the pattern of numbers on the cards below.



If the pattern continues, what number will be on the next card?

- A 90
- B 93
- C 95
- D 97

**Concurrent reporting notes page**

#5

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

**Other noteworthy comments:**

5

Each person running in a race will get a card with a number on it. Look at the pattern of numbers on the cards below.



If the pattern continues, what number will be on the next card?

- A 90
- B 93
- C 95
- D 97

Now I'm going to ask you some questions about the question you just answered.

### Retrospective Questions

1. Do you think you got this question correct? Yes or No
  - a. Why or why not?
  
2. Was this question hard, in the middle or easy for you?
  - a. Why?
  
3. Is this something you have learned in your math class before?  
Yes or No or I don't know (*or indicate other response*)
  
4. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
5. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

Great, thank you, now I'd like you to answer another question.

**6**

**Look at the pattern of numbers below.**

75, 77, 80, 84, 89, —?

**The pattern continues in the same way. What will be the next number in the pattern?**

**A** 90

**B** 93

**C** 95

**D** 97

**Concurrent reporting notes page**

#6

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

**Other noteworthy comments:**

6

**Look at the pattern of numbers below.**

75, 77, 80, 84, 89, —?

**The pattern continues in the same way. What will be the next number in the pattern?**

A 90

B 93

C 95

D 97



Now I'm going to ask you some questions about the two questions you just answered that were similar.

### Item Comparison Questions

1. Which question lets you know what you are supposed to do? First, second, or no preference?
  - a. Why?
  
2. Which of the two questions do you like better? First, second, or no preference?
  - a. Why?
  
3. Do you think that these two questions are asking you about the same math? Yes or No
  - a. If Yes, then ask for both items at once,
    - i. If No, then ask for each item version specifically:
      - ii. If I were a student in your class and I did not understand what this question was asking, could you explain what the first question is asking me to do in your own words?
      - iii. If I were a student in your class and I did not understand what this question was asking, could you explain what the second question is asking me to do in your own words?

Okay, let's move onto the next question.

7

**At a school, there are 704 desks to place into 22 classrooms. If the same number of desks is placed in each classroom, how many desks will be in each room?**

**A** 32

**B** 34

**C** 42

**D** 44

**Concurrent reporting notes page**

#7

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

**Other noteworthy comments:**

7

**At a school, there are 704 desks to place into 22 classrooms. If the same number of desks is placed in each classroom, how many desks will be in each room?**

**A** 32

**B** 34

**C** 42

**D** 44

Now I'm going to ask you some questions about the question you just answered.

### Retrospective Questions

1. Do you think you got this question correct? Yes or No
  - a. Why or why not?
  
2. Was this question hard, in the middle or easy for you?
  - a. Why?
  
3. Is this something you have learned in your math class before?  
Yes or No or I don't know (*or indicate other response*)
  
4. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
5. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

Great, thank you, now I'd like you to answer the last question.

8

**A school has 704 desks and 22 classrooms. Each classroom has the same number of desks. How many desks are in each classroom?**

**A** 32

**B** 34

**C** 42

**D** 44

**Concurrent reporting notes page**

#8

**ITEM CONTENT** Take notes on student's ability to answer correctly. Include any information on any noticeable errors in the math, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Is the student answering correctly? Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between.

**ITEM FORMAT** Take notes on any references to the format of the item. This includes making mention or using a picture to solve the question or any confusion with the layout of an item or information in a picture.

**ANSWERING STRATEGY** Summarize how student solved the problem, the steps that he or she is taking. Make a note for the 2<sup>nd</sup> item whether the strategy is the same or different as the 1<sup>st</sup> item presented.

**Other noteworthy comments:**

8

**A school has 704 desks and 22 classrooms. Each classroom has the same number of desks. How many desks are in each classroom?**

**A** 32

**B** 34

**C** 42

**D** 44

Now I'm going to ask you some questions about the question you just answered.

### Retrospective Questions

1. Do you think you got this question correct? Yes or No
  - a. Why or why not?
  
2. Was this question hard, in the middle or easy for you?
  - a. Why?
  
3. Is this something you have learned in your math class before?  
Yes or No or I don't know (*or indicate other response*)
  
4. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
5. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

Now I'm going to ask you some questions about the two questions you just answered that were similar.

### Item Comparison Questions

1. Which question lets you know what you are supposed to do? First, second, or no preference?
  - a. Why?
  
2. Which of the two questions do you like better? First, second, or no preference?
  - a. Why?
  
3. Do you think that these two questions are asking you about the same math? Yes or No
  - a. If Yes, then ask for both items at once,
    - i. If No, then ask for each item version specifically:
      - ii. If I were a student in your class and I did not understand what this question was asking, could you explain what the first question is asking me to do in your own words?
      - iii. If I were a student in your class and I did not understand what this question was asking, could you explain what the second question is asking me to do in your own words?

Thank you very much! That is the end of the test today. I really appreciate all your hard work. You can go back to class now.



Stop recording and set up for next student if applicable.

Appendix B  
Student Responses

	Student ID					
	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08
Date of interview	June 2, 2010	June 2, 2010	June 2, 2010	June 2, 2010	June 2, 2010	June 2, 2010
ELL status	NES	NES	NES	NES	NES	NES
Grade level	6	6	6	6	6	6
Test subject	Math	Math	Math	Math	Math	Math
Item #	5M	6M	6O	7M	7O	8M
Item text	Yesterday, there were 2,467 people at the museum. Today, there are 310 more people at the museum. How many people are at the museum today?	Last year, there were exactly 2,467 students attending Lee Elementary School. This year, there are 310 more students attending the school. How many students are attending Lee Elementary School this year?	Each small square on the grid is 1 square unit. How many total square units are shaded?	Each small square on the grid is 1 square unit. How many square units are needed to make the shaded figure shown on the grid?	Lisa rented 4 videotapes for \$4.80. How much did each tape cost to rent?	Sarah tossed a coin 10 times. She got heads 4 times and tails 6 times. Which chart shows these tosses?
Test booklet item #	2	4	3	6	5	8
Test form	2	2	2	2	2	2
Respondent's answer	C	C	C	D	D	B
Issues with item content	None	None	None	No confusion (very confident), but multiplied rather than dividing.	No confusion (very confident), but multiplied rather than dividing.	None
Issues with item format	None	None	None	None	None	None
Answering strategy	Knows she had 2,467 yesterday, 310 more today. Says "more," so needs to add. Writes out addition problem, explains it aloud.	Knows last year there were 2,467, this year 310 more than that, so would have to add them together to see the answer. Writes out addition problem, explains it aloud.	Draws lines where grid lines would be in shaded area; counts squares in shaded area.	Draws lines where grid lines would be in shaded area; counts squares in shaded area.	Knows she rented 4 videotapes at the same price of \$4.80. So add \$4.80 4 times, or times it by 4. Easier to times by 4, writes out multiplication problem, explains it aloud.	Sees a chart that shows heads and gives the word, and shows tails and gives the word. Goes through each answer choice and how many times each lands on head/tail, knows that heads has to be for 4 times, tails for 6 times.

Appendix: Continued

	Student ID							
	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	
Other noteworthy observations (include Spanish info here)								
Do you think you got this question correct?	Yes—had to add, lined up numbers correctly.	Yes—put the addition sentence in proper order-lined digits up.	Yes—the key says a square on the grid is 1 square unit. Says “how many” so knows she has to count.	Yes—made the lines that it shows her. Used the key—said the squares on the grid is one square unit.	Yes—They tell you that she buys 4 cookies for \$4.80, so times \$4.80 by 4.	Yes—Automatically knew she had to add \$4.80 four times, or times it by 4.	Yes—It tells how many times it lands on heads and tails. Gives pictures to show, and gives a key to know which one is which.	Yes—Knows which looks like head and tail so it helps her, and it tells her how many times it landed on heads and tails.
Was this question hard, in the middle, or easy for you?	Easy—They gave good explanations.	Easy—It’s easy for her to add low numbers.	Easy	Easy—just have to count the squares, and they gave a key.	Yes—gave a clue word, “how much.”	Easy—Multiplication is easiest one to do, adding would have been harder.	Easy—gives clue words—says “which chart shows these tosses,” knows she has to look at charts. Key shows which is which when she looks at the charts.	Easy—It gave a clue word and clue pictures. In the clue pictures it uses tally marks to see how much, knows each tally mark represents 1.
Is this something you have learned in your class before?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Were there any words, ideas, or anything else that made it easy to answer this question?	more yesterday, today,	“last year” and the exact numbers in the school; “this year”—knew she had to add.	Yes—the key, “how many”	Yes—the key; says “how many”—knew she had to count	Yes—Knew she had to times or add.	Yes—gave how much she rented, “how much”—knew she’d have to add or times.	Yes—the key with heads and tails, which side they landed on, how many times.	Yes—they give you that the coin landed on heads four times, on tails six times. They give you which side, how many times, so you can see how many tallies.
Were there any words, ideas, or anything else that made this question harder or confusing to answer?	No	No	No	No	No	No	No	No

Appendix: Continued

	Student ID			
	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08	JL-N4-M-91-08
Which question lets you know what you are supposed to do? First, second, or no preference? Why?	No preference — both numbers to add the same, adding, same answer.	No preference. Both have the same thing, key, and you need to count.	No preference. Same numbers, both say “how much.”	No preference. Both have the same question, and the same “how much times.”
Which of the two questions do you like better? First, second, or no preference? Why?	No preference. Both the same question, just not at the same place.	No preference. Both give the same question. Just use two different words but really the same meaning.	No preference. Both have the same idea of timesing or adding, because they both have same numbers. Both have same numbers.	No preference. They’re both about the same thing, things you have to add, both have keys.
Do you think that these two questions are asking you about the same math/science? Yes or No	No (based on previous answer, think she just misunderstood this question)	Yes	Yes	Yes
If I were a student in your class and I did not understand what this question was asking, could you explain in your own words what the first question is asking me to do?	This question is asking you to add — use a clue word. They say (reads question).	Rereads question. “one square unit is one square on the grid. How many means add or count.”	Questions are asking (rereads second question). They say “how much,” so you can either add \$4.80 four times, or times \$4.80 four times. Explains multiplication problem.	The question says that (rereads second question). First know that it lands on heads four times, and on the key you would look at heads. You know that it landed on tails six times, and on the key look at tails and it shows you. So you’d have to look at all your charts before you answer. Heads has to have 4, and tails has to have 6, and it’s using tally marks.
If I were a student in your class and I did not understand what this question was asking, could you explain in your own words what the first question is asking me to do?	They’re saying (reads question) ... gives you the clue word “more” so you know you have to add, they say “how many people.”			

### Appendix C

#### Student Performance and Responses to Interview Questions by Grade and Subject

The data for each respective grade and subject was divided into four groups, each one corresponding to the item pair’s score pattern: (a) both items were answered correctly, (b) the first presented item was answered correctly and the second item was answered incorrectly, (c) the first item was answered incorrectly and the second correctly, and (d) both items were answered incorrectly (see the first tables in each section).

#### Grade 4 Mathematics

Table C1 Score Patterns for Grade 4 Mathematics Item Pairs

Score pattern	Order of item presentation			
	Original → Modified		Modified → Original	
	English language learner	Non-English language learner	English language learner	Non-English language learner
Both correct	27	24	27	30
First correct, second incorrect	8	0	2	0
First incorrect, second correct	3	0	5	2
Both incorrect	11	4	10	4

The results for each respective grade and subject focus on the middle two score patterns for the ELL student response only. To help the discussion, the results are divided into the features of the test items that helped and those that hindered students in solving the items. The modified items, presented first or last, that were answered correctly are identified as items that helped students in solving the items, while the modified items, in either position, that were answered incorrectly are identified as items that hindered students in solving the items.

#### Features That Helped

Table C2 English Language Learner Students’ Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	3	3
Medium	2	1
Hard	0	1

When referring to a modified item, three students stated that it was an easy item because the math they were asked to complete was easy for them (e.g., “multiplication is easy”). Three students stated an original item was easy because they had already learned the material. One of the students believed the modified material was difficult and one thought an original item was difficult because they “never saw question before.”

Table C3 Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
3	1

Two of the three students stated phrases like “how many more” or “total number of balloons” made the question easier. One student suggested the picture made the modified question easier.

Table C4 Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
3	2

Three students stated that some words (e.g., tally, tails, tossed, and landed) in the original item made the “question more confusing.”

**Table C5** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
1	1	3

One student stated that the modified item was “easier.”

**Table C6** Which Question Do You Like Better?

Modified	Original	No preference
1	4	0

A student stated that the modified item “gave more example” and therefore was easier. The students who preferred the original items felt they were easier.

**Table C7** Do You Think That These Two Questions Are Asking About the Same Science?

Yes	No
4	1

### Features That Hindered

**Table C8** English Language Learner Students’ Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	6	5
Medium	3	1
Hard	2	5

Two students thought the modified items were difficult because they “didn’t understand much.” Five students thought that the original items were difficult because they didn’t understand the questions (e.g., “didn’t understand what to do at first,” “question didn’t say if you should add, subtract, or multiply”).

**Table C9** Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
4	4

Four students stated that a phrase in the original item (e.g., “what was the total amount,” “how much will 4 pounds cost”) made the question easy to answer. Four students stated that a part of the modified item (e.g., “fraction,” “how many total”) made it easy to answer.

**Table C10** Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
1	5

Five students stated that some words (e.g., “all together, how many,” “What’s the value of 6”) in the original item made the item harder or confusing to answer. A student stated “how much” made the modified item harder to answer.

**Table C11** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
2	3	7

Two students preferred the modified item because it explained more. Three students preferred the original item because of similar reasons.

**Table C12** Which Question Do You Like Better?

Modified	Original	No preference
4	3	5

Four students preferred the modified items because they gave more information or better explanations. While three students stated they preferred the original items.

**Table C13** Do You Think That These Two Questions Are Asking About the Same Science?

Yes	No
10	2

### Grade 4 Science

**Table C14** Score Patterns for Grade 4 Science Item Pairs

Score pattern	Order of item presentation			
	Original → Modified		Modified → Original	
	English language learner	Non-English language learner	English language learner	Non-English language learner
Both correct	16	28	22	20
First correct, second incorrect	10	3	10	4
First incorrect, second correct	8	4	3	6
Both incorrect	16	8	11	10

### Features That Helped

**Table C15** English Language Learner Students' Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	12	6
Medium	15	7
Hard	1	5

When referring to a modified item, a student stated that it was a somewhat difficult (i.e., medium difficulty) item because the student “understood some words,” but when referring to the original item, the student stated that it was a hard item because the student didn’t “understand axe is wide at one end and pointed at the other.”

**Table C16** Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
12	9

Three students responded that the picture/figure made this question easy to answer. Other students stated the change in how the question was worded (e.g., “These three plants grew next to a sunny window for 6 weeks. The reason they look different is they had—” instead of “These three plants have all been growing on the same sunny windowsill for 6 weeks. What may have caused the difference in appearance?”) made the question easier.

**Table C17** Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
7	10

Six students stated that some words (e.g., data, observation, characteristics, loggerhead) in the original item made the question harder or confusing.

**Table C18** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
15	2	1

Ten students stated that the modified items were easier or explained better.

**Table C19** Which Question Do You Like Better?

Modified	Original	No preference
15	2	1

Students stated that the modified items were “easier” and better understood and therefore could “answer better.”

**Table C20** Do You Think That These Two Questions Are Asking About the Same Science?

Yes	No
10	8

### Features That Hindered

**Table C21** English Language Learner Students’ Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	6	6
Medium	7	5
Hard	0	2

None of the students thought that the modified items were difficult. Two students thought that it was hard/difficult to “understand” the words or answer choices in the original items.

**Table C22** Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
10	9

Four students stated that the picture in the modified version made the question easy to answer. Another four said a phrase in the question (e.g., “breaks into pieces”) made the modified questions easy. Four students stated that the picture in the original version made the question easy to answer. Four students stated that some part of the original item (e.g., “Tiny turtles hatch from the eggs and find their way to the ocean water”) made the question easy to answer.

**Table C23** Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
8	5

Eight students stated that some words (e.g., evaporation, fur) in the modified item made the item harder or confusing to answer. Four students thought a word or phrase (e.g., three plants on the same windowsill but were different sizes) made the original item harder to answer.

**Table C24** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
5	7	1

Seven students stated that the original item gave “more details” and “more instructions,” while four students stated that the pictures for the modified questions were helpful and the questions gave “more clues.”

**Table C25** Which Question Do You Like Better?

Modified	Original	No preference
7	6	1

Five students stated that the original items were “easier” and gave “more information,” while five students stated that the modified item “gave more clues” or “had less difficult words.”

**Table C26** Do You Think That These Two Questions Are Asking About the Same Science?

Yes	No
9	4

## Grade 6 Mathematics

**Table C27** Score Patterns for Item Pairs

Score pattern	Order of item presentation			
	Original → Modified		Modified → Original	
	English language learner	Non-English language learner	English language learner	Non-English language learner
Both correct	14	22	9	30
First correct, second incorrect	6	5	6	6
First incorrect, second correct	6	7	6	4
Both incorrect	9	7	13	7

## Features That Helped

**Table C28** English Language Learner Students’ Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	8	5
Medium	1	3
Hard	1	1

A few more students thought the modified items were easier. Students who preferred the modified items provided responses such as “it’s the same as before, except this time I have to use pencils,” recognizing the modified and original items are asking the same math, or they realized what they had to do after re-reading the question once. Students indicated that the original items were easy because it was obvious or they understood everything. One student thought the original item was hard because he didn’t understand the question.

**Table C29** Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
5	5

One student indicated that for both the modified and the original items that “how many” was helpful because it indicated addition. Two students who identified that something helped in the modified items said everything helped. Students who indicated something helped in the original items identified one phrase in the stem that was crucial for understanding, such as “how many desks are in each classroom” or “continue the pattern.”

**Table C30** Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
7	2

More students found words or components of the original items hard or confusing as compared to the modified items. One student stated in general the language was difficult in the original item, one student identified a specific word (spinner), and three students indicated specific phrases in the original items. Both cases of the hard or confusing information in the modified items were referring to a phrase in the stem.

**Table C31** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
5	3	12

Most students did not have a preference; only two students provided a reason for no preference, indicating they were the same question in one case and that both items explained what you are supposed to do well in the other case. The reasons for preferring the modified were that it was clearer, easier, or helps. The explanations for preferring the original item were similar with students indicating the original item was simple or explained better what you need to do.

**Table C32** Which Question Do You Like Better?

Modified	Original	No preference
5	4	11

Most students did not have a preference and only one student provided a rationale, indicating that both showed drawings that were helpful. Students who preferred the original item mentioned that it better explained what the student should do. One student indicated he preferred the original because it had a better scenario.

**Table C33** Do You Think That These Two Questions Are Asking About the Same Math?

Yes	No
7	3

### Features That Hindered

**Table C34** English Language Learner Students’ Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	4	4
Medium	3	3
Hard	4	4

The level of difficulty as perceived by the students followed the same pattern for the modified and original items. Students thought the modified items were easy because they recognized the items were the same, or because they understood the items. Students thought the modified items were medium because they didn’t understand many of the words. One student who thought a modified item was in the middle of difficulty misinterpreted an item. The item said “the pattern repeats the first six shapes in the same order” and the student thought the question was asking about the first and sixth shapes. Students indicated items were hard when they did not understand the question; the numbers in the question; or the meaning of a specific word, “closest.” Students thought the original items were easy because they recognized the patterns or thought parts of the question gave a hint towards the answer. They thought the original items were in the middle when they were unsure of their answer, and students thought the items were hard when they didn’t understand components of the items, such as how a pattern repeated or the amount of a liter.

**Table C35** Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
5	6

Five students found something in the modified items that made it easier to answer. One student indicated the numbers below the item stem helped, and three students mentioned specific phrases, such as “the first six shapes” and “find the chart.” One student said everything made it easier to answer the modified. Six students found something in the original item easier to answer. Five of the six students mentioned one or more specific words such as “equation,” “rule,” and “continues.” One student said “specific words and details” without describing the words or details. That same student thought the scenario helped. Lastly, one student found the picture very helpful.

**Table C36** Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
6	6

There were six cases of students finding something hard in a modified item and six cases of students finding something hard or confusing in original items. Words and phrases in the modified items included “perimeter,” “liter and pint,” “closest,” and “without looking.” One student did not understand the question. Similar issues were the reasons why students found the original items confusing. Words included “liter, pint, and ounce,” “chicken broth,” and “finding.” One student didn’t understand the entire first sentence, and another thought the “English, and the pattern was too confusing to understand.”

**Table C37** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
2	3	6

The two students who preferred the modified items indicated that they understood it more. The three students who thought the original item let them know what they were supposed to do said that they like the original because it gave an example; because it looked better; and because the original item, which was the first of the pair, gave the student an

idea of what to put in the next question. Four of the students who stated they did not have a preference of items gave an explanation. Three of the four students said that the two questions asked the same or nearly the same thing, and one student did not understand either item.

**Table C38** Which Question Do You Like Better?

Modified	Original	No preference
2	5	3

Two students preferred the modified item, saying that it was clear and that the modified item “explained more which helped me understand.” Five students liked the original item more and provided reasons including that it “shows a picture and explains what to do better,” “explains more about the problem and I like this scenario better,” “it gives me a whole sentence,” “the words help me a little bit,” and “it tells you what to do.” One student also preferred the original because it was the first of the pair of items and seeing the first item “gave me an influence of what to put in the next question.” Two students explained why they did not have a preference for the modified or the original. One student thought they demonstrated the same thing, and the other student did not understand either question.

**Table C39** Do You Think That These Two Questions Are Asking About the Same Math?

Yes	No
8	3

### Grade 6 Science

**Table C40** Score Patterns for Item Pairs

Score pattern	Order of item presentation			
	Original → Modified		Modified → Original	
	English language learner	Non-English language learner	English language learner	Non-English language learner
Both correct	11	21	10	20
First correct, second incorrect	5	5	8	6
First incorrect, second correct	6	5	5	3
Both incorrect	14	13	13	12

### Features That Helped

**Table C41** English Language Learner Students’ Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	2	2
Medium	7	9
Hard	4	4

When referring to a modified item, a student stated that it was a somewhat difficult (i.e., medium difficulty) item because the question made the student “think hard,” but when referring to the original item, the student stated that it was a hard item because the student “didn’t understand anything about the question.”

**Table C42** Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
9	5

Six students responded that the picture/figure made this question easy to answer and one student stated that the modified item was a “very clear question.” Other students stated that simplifying the vocabulary (e.g., using *insect* instead of *lacewing*, using *giving* instead of *providing*) made the question easier.

**Table C43** Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
8	9

Five students stated that some words (e.g., *lacewing*, Isaac Newton) in the original item made the “question more confusing.”

**Table C44** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
8	1	4

Three students stated that the modified items were in an “easier form” and these students were able to “understand more.” The items were modified by refining the context, simplifying the vocabulary, making the stem concise, and showing a graphic representation.

**Table C45** Which Question Do You Like Better?

Modified	Original	No preference
8	3	3

Students stated that the modified items were “easier” and that they “could understand [them] better.” On the other hand, those who preferred the original items liked their “challenge” and thought that the “answers went better with the stem[s].”

**Table C46** Do You Think That These Two Questions Are Asking About the Same Science?

Yes	No
10	4

### Features That Hindered

**Table C47** English Language Learner Students’ Rating of Item Difficulty

Item difficulty	Modified	Original
Easy	3	5
Medium	6	3
Hard	0	1

Four students did not, or thought that it was hard/difficult to, “understand” the words/graph in the items modified with refined context, concise stems, and concise options. Two students thought that the original items had words that were “difficult.” One student had difficulty reading the original item.

**Table C48** Yes, There Were Words, Ideas, or Anything Else That Made This Question Easy to Answer

Modified	Original
2	4

Two students stated that the picture in both the modified and original versions made the questions easy to answer. Two students stated that some part of the original item (e.g., last sentence) made the question easy to answer; their modified counterpart items were modified with refined context, simplified vocabulary, concise stems, and graphic representations. One student stated that the item that was modified by refining its context made it easy to answer it.

**Table C49** Yes, There Were Words, Ideas, or Anything Else That Made This Question Harder or Confusing

Modified	Original
2	4

Two students stated that some words (e.g., lacewing, providing) in the original items made the items harder or confusing to answer, whereas their modified counterpart items (i.e., simplified vocabulary, refined context, concise stems, and/or graphic representations) were not harder or confusing to answer. One student stated that the simplified verb form version of the item was “confusing at first.”

**Table C50** Which Question Lets You Know What You Are Supposed to Do?

Modified	Original	No preference
3	3	3

Two students stated that the original items were “easier” and gave “more information,” while one student stated that the modified item was “more direct” (item was modified for refined context, and its stem and choices were concise).

**Table C51** Which Question Do You Like Better?

Modified	Original	No preference
3	5	1

Four students stated that the original items were “easier” and gave “more information,” while two students stated that the modified items were either “more direct” or “easier” (items were modified for refined context, and their stem and choices were concise).

**Table C52** Do You Think That These Two Questions Are Asking About the Same Science?

Yes	No
8	1

### Suggested citation:

Young, J. W., King, T. C., Hauck, M. C., Ginsburgh, M., Kotloff, L., Cabrera, J., & Cavalie, C. (2014). *Improving content assessment for English language learners: Studies of the linguistic modification of test items* (Research Report No. RR-14-23). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12023

**Action Editor:** John Sabatini

**Reviewers:** Danielle Guzman-Orth and Sultan Turkan

ETS, the ETS logo, and LISTENING. LEARNING. LEADING. are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at <http://search.ets.org/researcher/>