

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications: Department of Teaching, Learning and Teacher Education Department of Teaching, Learning and Teacher Education

2014

Equitable Written Assessments for English Language Learners: How Scaffolding Helps

Marcelle A. Siegel

Deepika Menon

Somnath Sinha

Nattida Promyod

Cathy Wissehr

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/teachlearnfacpub>



Part of the [Curriculum and Instruction Commons](#), [Language and Literacy Education Commons](#), and the [Teacher Education and Professional Development Commons](#)

This Article is brought to you for free and open access by the Department of Teaching, Learning and Teacher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications: Department of Teaching, Learning and Teacher Education by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Marcelle A. Siegel, Deepika Menon, Somnath Sinha, Nattida Promyod, Cathy Wissehr, and Kristy L. Halvorson

Published in *Journal of Science Teacher Education* 25:6 (2014), pp. 681–708; doi: 10.1007/s10972-014-9392-1
Copyright © 2014 The Association for Science Teacher Education, USA. Published by Routledge/Taylor
& Francis Group. Used by permission.
Published online September 25, 2014.

Equitable Written Assessments for English Language Learners: How Scaffolding Helps

Marcelle A. Siegel,¹ Deepika Menon,¹ Somnath Sinha,¹ Nattida Promyod,²
Cathy Wissehr,³ and Kristy L. Halverson⁴

1. MU Science Education Center, University of Missouri, Columbia, Missouri, USA
2. University of Iowa, Iowa City, Iowa, USA
3. University of Arkansas, Fayetteville, Arkansas, USA
4. University of Southern Mississippi, Hattiesburg, Mississippi, USA

Corresponding author – Marcelle A. Siegel, MU Science Education Center, University of Missouri, 303 Townsend Hall, Columbia, MO 65211, USA, email siegelm@missouri.edu

Abstract

This study investigated the effects of the use of scaffolds in written classroom assessments through the voices of both native English speakers and English language learners from two middle schools. Students responded to assessment tasks in writing, by speaking aloud using think aloud protocols, and by reflecting in a post-assessment interview. The classroom assessment tasks were designed to engage students in scientific sense making and multifaceted language use, as recommended by the Next Generation Science Standards. Data analyses showed that both groups benefited from the use of scaffolds. The findings revealed specific ways that modifications were supportive in helping students to comprehend, visualize and organize thinking, and elicit responses. This study offers a model for both sensitizing teachers and strengthening their strategies for scaffolding assessments equitably.

Keywords: assessment, English language learner, scaffolding

Introduction

Language and culture influence how students approach situations and construct knowledge, making the need for new forms of equitable assessments for English language learners (ELLs) clear (Lee, 2001; Solano-Flores & Nelson-Barber, 2001). Developers of content assessments—from researchers, to curriculum developers, to teachers—rarely have the expertise to take language and culture into account. For example, most teachers report that they are not equipped to teach ELLs while teaching content (e.g., National Center for Education Statistics [NCES], 2010). In addition, traditional assessments are not helpful in accurately measuring content knowledge of ELLs (Jenelle, 2004). Yet, recent research has shown that well-designed assessments can be equitable for ELLs (e.g., Abedi, Hofstetter, & Lord, 2004; Siegel, 2007). Such assessments should provide opportunities for diverse students to demonstrate their knowledge concurrent with their backgrounds (Lee, 2005).

Nevertheless, few assessments are available that offer such opportunities. Studies show that current classroom assessments and standardized tests are not equitable for all groups of students (Abedi, 2004; Solano-Flores, 2008; Solano-Flores & Nelson-Barber, 2001). Separating language and content is very challenging and complex because language and content are so interlinked. This small-scale study attempts to tease these complex links between language and content out by examining assessments that were designed for ELLs in middle school life science. The assessment tasks address several of the Disciplinary Core Ideas, Crosscutting Concepts, and Science and Engineering Practices outlined in the Next Generation Science Standards (National Research Council [NRC], 2013). We compare modified life science assessment tasks specifically designed for ELLs with the original assessment tasks. Our data sources included student work on the assessment tasks, interview data, and think aloud protocols of both ELLs and native English speakers (NES). We discuss ways that particular forms of scaffolding appeared to assist ELLs in comprehending, reasoning about, and responding to the written assessment tasks. Through the voices of ELLs, this article provides teachers a glimpse into how students reason through content assessments and the specific differences that scaffolds afford.

Teachers and Equitable Assessment

Assessment is a powerful tool in the classroom because it can uncover inequitable conditions and facilitate interaction between teacher and students about learning (Lyon, 2013). While the research on equitable assessment is limited, based on the slightly more extensive literature involving teachers and ELLs, we know that several factors inhibit teachers from using equitable classroom assessments. One barrier is that not all teachers think it is their responsibility to use equitable strategies to teach diverse students (Bryan & Atwater, 2002; Lee, Luykx, Buxton, & Shaver, 2007). In addition, research shows that teachers receive little preservice preparation or professional development on equitable assessment (Hazel, Logan, & Gallagher, 1997; Lyon, 2013; Wong-Fillmore, 2007). Another set of barriers that limit teachers' time and motivation are pressure from highstakes testing, unclear testing policies, and discriminatory policies (such as "English only") (Hazel et al., 1997; Lee et al., 2007; Wiley & Wright, 2004).

Three barriers to equitable assessment are particularly important to this study. One aspect of effective assessment practice is teachers' beliefs and values (Abell & Siegel, 2011). Preservice and inservice teachers may hold deficit beliefs about diverse students, including ELLs, such as not understanding cultural influences on learning or viewing language minorities as possessing low intelligence or ability. Most classroom teachers have concerns about their preparation to support ELLs' learning (Lewis-Moreno, 2007; NCES, 2010). A second factor that limits teacher effectiveness is resources. Few equitable assessment materials are available. Moreover, there is a lack of culturally relevant curriculum materials available for science teachers (LaCelle-Peterson & Rivera, 1994; Lee et al., 2007). Third, teachers might be familiar with "scaffolding" or quality assessments but do not understand the rationale behind why the tools are effective (Hobsbaum, Peters, & Sylva, 1996). What is it about a teaching technique that assists ELLs to complete a science activity, or more specifically, what is it about a particular scaffold that helps ELLs better comprehend an assessment?

Modifications and Scaffolds

Complex assessment tasks, or performance assessments, are central to this study. These types of written science assessments provide opportunities for ELLs to learn and demonstrate science understanding and expertise and create spaces for them to develop further skills in English (Shaw, Bunch, & Geaney, 2010). Modifications come into play because the performance of ELLs, or any student, has been found to be very dependent on the wording of assessment prompts (Baxter, Shavelson, Goldman, & Pine, 1992). One strategy for increasing equity is accommodation, or in this case modifications, that change the written assessment (while accommodations might alter the way it is administered). For example, written modifications could include adding visual supports, reducing linguistic complexity, offering customized glossaries, and revising the directions. Research on the assessment of ELLs consistently points to unnecessary linguistic complexity of assessments as an inequity that does not give ELLs the opportunity to show what they know (Abedi, 2009).

Modifications, such as reduction of extraneous linguistic complexity, can be viewed as a form of scaffolding. Scaffolding is the support that is contingent on the needs of the learner. Wood, Bruner, and Ross (1976) introduced scaffolding as tutorial interactions that a knowledgeable adult provides for a child. Some argue that scaffolding can occur only within the zone of proximal development (Vygotsky, 1978). This zone refers to the area between a student's independent thinking and the level of thinking with the guidance of people or tools. The learner is thus assisted to perform at a higher level than one could accomplish alone. With scaffolding, learners can achieve more advanced skills and reach levels of advanced cognition.

We argue that scaffolding includes the use of written modifications and not just interaction with a skilled tutor. The view of Vygotsky, who discussed the use of artifacts and tools in assisting learning (1978), aligns with this definition. Palincsar (1998) also supports our perspective that researchers attribute not only interaction between individuals as scaffolding but also the interactions that occur between individuals and artifacts, and that the zone of proximal development is certainly embedded in particular contexts.

Our equity framework for modifying written assessments was developed to support, challenge, and enhance learning of all students based on five principles (Siegel, 2007). The principles for adapting assessments for ELLs included: (1) matching the learning goals and language of instruction, (2) ensuring cultural and linguistic comprehensibility, (3) challenging students to think about difficult ideas on assessments, (4) eliciting students' understanding, and (5) scaffolding the use of language in assessments to support student learning. Siegel, Wissehr, and Halverson (2008) explain that basically classroom assessments for ELLs should be comprehensible, challenging, and supportive of learning. To address the first principle, for example, a teacher using the term "investigation" during instruction would also use this term on the assessment, rather than a synonym such as "experimentation." To be comprehensible, a teacher can take a passage that may cause extra reading time for an ELL and reduce the text (Abedi, Lord, Hofstetter, & Baker, 2000; CRESST, 2001). The second principle also involves cultural comprehensibility, for instance changing the context of a question that might be unfamiliar to a cultural group to a more familiar context. The third principle ensures that reducing the complexities involved with the language should not reduce the opportunities for students to think and reason critically about high-level concepts. This is an important guideline because teachers might reduce the cognitive complexity of an assessment for ELLs and thus diminish opportunity to learn (Walqui, 2003). The fourth principle relates to the need for students to be able to demonstrate their understanding through their written responses. If students are not responding to a prompt or answering off-topic, the assessment likely requires revision. Finally, the fifth principle supports the idea of using scaffolds, such as sentence starters, graphic organizers, and additional prompts, to help students meet the cognitive demands of the assessment.

Research Questions

This study focused on two versions of assessment tasks, the original version and the modified version. Both ELLs and NESs completed the assessment tasks. In general, we wanted to know how effective both types of assessment tasks were for students. We did not want to assume that the modified tasks would be more effective, so we asked first, "Do the modified assessments versus original assessments support or impede ELLs and NESs?" For the modified tasks, we sought to understand the ways the scaffolding actually affects students, from the student's perspective. We examined, "How does scaffolding in the modified assessment tasks support ELLs (and NESs)?"

Methods

Researchers

The researchers' main field is science education. We are a group of science education researchers, teachers, and teacher educators; half of us are faculty and half graduate students. We have expertise in assessment, science teacher professional development, and learning with/through visual representations. It is important to note that we are not language experts, although two of the authors have previous background with teaching and assessing ELLs. In addition, three of us are L1 English and three of us are L1 Thai, Bengali, or Hindi.

Participants

The participants of the study included six students from two middle schools in the US. For this small group, we aimed to select a mix of ELLs and NESs and include both males and females. The criteria for selection of ELL participants included same first language spoken (Spanish) and advanced English language status. Because the modified *written* assessments were designed for a high level of English ability (in contrast to, for instance, *verbal* modifications for those *newer* to English), we aimed to find students who were advanced ELLs as identified and recommended by teachers. Unfortunately we do not report language level because the state in which the study took place does not use a sophisticated language test for reporting a particular level of fluency. Among the six participants (see Table 1), two were females and four students were males; three were Latina/o ELLs and three were NESs.

Student	Grade	Gender	Ethnicity	L1	ELL status
Antonia	7th	F	Latina	Spanish	Yes
Fernando	7th	M	Latino	Spanish	Yes
Joaquin	7th	M	Latino	Spanish	Yes
Dave	7th	M	Caucasian	English	No
Chloe	7th	F	Caucasian	English	No
Owen	7th	M	Caucasian	English	No

The students were asked to read the question and explain their thinking regarding the question simultaneously (think aloud protocol). The students were encouraged to answer the question as best as possible, followed by post-assessment interview questions. The data sources (Table 2) included the students' work on the assessment tasks, transcripts of the audiotaped think aloud protocols, and transcripts of the post-assessment interviews—each discussed in the following sections.

Source	Data
Assessment tasks	Written student work
Think aloud protocol	Transcript of audiotaped student comments
Post-assessment interviews	Transcript of audiotaped student comments

Assessment Tasks

Researchers collected written student work on the assessment tasks. Here we describe the development of these key parts of the study. Some of the assessments were developed by a curriculum team with funding from the National Science Foundation and have been previously modified and studied (Siegel, 2007; Siegel, Markey, & Swann, 2005). In addition, new assessments were designed for this study. The assessments were completed before the Next Generation Science Standards (NGSS; NRC, 2013) and frameworks were released. However, as the NGSS drafts became available we attempted to align the assessments with

the new standards. The assessment tasks address two Disciplinary Core Ideas in middle school life science: MS-LS1 from Molecules to Organisms Structures and Processes and MS-LS2 Ecosystems Interactions, Energy, and Dynamics. The assessment tasks target several Crosscutting Concepts and Scientific and Engineering Practices of the recently released standards as well. For further details, each NGSS link is stated next to the question number in Appendix 1.

Twelve assessment tasks were developed and pilot tested using an equity framework consisting of research-based principles, called McCes (Siegel, 2007; Siegel et al., 2008). Six “original” assessments were revised for advanced ELLs without reducing the science content to form six “modified” assessments.

An example of an original and modified assessment is shown in Figure 1. In addition, all of the original and modified assessment tasks are available in Appendix 1 (Version A and Version B). Modifications of the items were intended to maintain the content while simplifying the language. The modifications (see Appendix 2) included:

- Simplified language terms
- Simplified grammar and sentence structure
- Addition of visuals, such as example diagrams or photographs
- Tables with space to enter information
- Division of question into parts
- Making relevant connections between various parts of the question
- Reduction of sentences and text.

Due to the length of the complex, written assessment tasks, the assessments were divided into two groups, A and B. Students responded either to group A or group B. Each group contained either an original or modified question (1, 2, 4, 5), and both groups contained modified question 3.



Zebra mussels clogging a pipe (1, 2).



Smaller zebra mussels attached to a larger, native mussel (1, 2).

- A. ~~Describe~~ What has been the impact of the zebra mussel on ~~Missouri ecosystems~~ ecosystems in Missouri in terms of ~~competition and the mussel's effects on man-made structures~~ the following: (3)
- Competition with other mussel species (4)
 - Effects on manmade structures and equipment (dams, boats, etc.) (4)
- B. How do you think these mussels may affect fisheries in the area?
- C. Considering what you know about how zebra mussels can spread, how would you plan to limit the spread of zebra mussel to new lakes and rivers? What steps might you suggest scientists take to begin permanently removing these mussels from the ecosystem?

Modifications

1. addition of visual support
2. brief note describing each visual support
3. reduction of words in the item stem
4. replacing a long question with two simple questions

Figure 1. Illustration of changes for one question.

Data Collection

We led assessments, think aloud protocols, and interviews with students during the normal school day. Researchers began by introducing the purpose of the session and conducting a practice think aloud problem with a participant. The introduction provided an opportunity for the student to become comfortable around the researcher. The student read the warm-up assessment aloud while practicing thinking aloud (Appendix 1). The process for collecting data included: (1) the student thinking aloud while responding to the assessment and (2) a post-assessment interview. One researcher audiorecorded each student during the data collection sessions. The combined sessions lasted up to 2 h. The sessions were conducted in English, although the students were allowed to draw and to use their first language and then prompted to explain also in English. Audiorecords were transcribed to serve as data sources.

Think Aloud Protocol

Thinking aloud is used as a cognitive tool to assess students' learning proficiency and to study problemsolving of students (e.g., Block, 1986). We asked students to tell us what they were thinking as they read through and tried to solve the assessment tasks. This involved a combination of speaking while thinking, beginning to write, voicing ideas, and completing the written response. Due to the time to complete one assessment, the original and modified assessments were divided randomly among participants so that they each responded to 3 original and 3 modified assessments. Block (1986, p. 464) explained that "think aloud protocols provide a direct view of a reader's mental activity, a kind of window into those processes which are usually hidden." For this study, think aloud protocols assisted researchers in examining assessment items to see how ELLs interpret them. The technique provided a glimpse into how ELLs approach an assessment task and what is useful or disadvantageous about the assessment task.

Post Assessment Interviews

After completing the assessment set while thinking aloud, the student was interviewed about what was most challenging and most helpful in the assessments and what they would revise to clarify the assessment. This post-assessment interview lasted about 25 min. The focus of the interviews was to determine the factors that affected the accessibility of answering these questions and how, according to the students, these difficulties might be alleviated.

Data Analysis

A group consisting of three researchers analyzed the data. The data from the written assessments, think aloud protocols, and post-assessment interviews were transcribed, and the criteria for coding the data were developed and listed as bulleted points by the team. Each assessment task had its own set of criteria based on the scaffolding in the question.

Examples of criteria include:

- Does the student understand the question in terms of:
 - flow of language,
 - vocabulary,
 - framing,
 - organization,
 - continuity,
 - connection between the sentences,
 - and comprehension of ideas?
- Does the student interpret the question in terms of intended outcome?
- Does the student make sense of the diagram?
- What parts of the question does the student find easy to answer? Why?

Researchers read and interpreted the data based on the criteria to analyze students' responses. Tables were used to summarize the analyses. One column listed a judgment about a criterion, such as "difficulty comprehending ideas," and another column listed quotes from the data as evidence. The three sources of data were interpreted for each analysis summary shown in a table, thus offering triangulation of data. The students were middle school level and were not expected to analyze every aspect of the assessments as a researcher would. There were many times when a student did not have a lot to say about an assessment. For example, when we noticed trends in the think aloud protocol, we did not expect the student to be able to reflect on this in the interview. In the next phase of analysis, the formulation of tables helped the researchers identify common themes emerging from the data based on the research questions. For example, did students find the same aspect of an assessment useful for a similar reason?

Trustworthiness was established through several techniques. We verified findings through triangulation of multiple data sources and multiple researchers (Lincoln & Guba, 1985). Brief peer debriefing with a fourth researcher on our team and language specialists outside our team was conducted to check our interpretations (Lincoln & Guba, 1985). In addition, we searched for negative cases that did not fit the emergent patterns of findings (Lincoln & Guba, 1985). Our post interview served as a form of checking with the student about his/her reactions to the assessments. While several strategies were used to establish trustworthiness, we also recognize that this study had limitations, such as a small number of participants, amount of tasks per interview period, and level of discourse analysis that could constrain the ability to generalize the findings.

Results

The following section reports findings in two parts, based on the research questions. The description of the results is supported by excerpts of the transcripts.

Research Question 1: Do the Modified Assessments Versus Original Assessments Support or Impede ELLs and NESs?

We found that the modified assessments supported both ELLs and NESs. Findings indicated that the ELLs and NESs were able to understand and answer the modified assessment items better than the original items. Modified question 4 was an example we will discuss the most. During the post interview, Owen (NES) stated that the organization of question 4 (modified version, see Appendix 1, version A) made it easier for him to answer. Owen stated:

I think the questions the easiest one for me to answer would probably be the cough syrup one. Because it was asking like, what you think, because it had graphs and it showed, I guess the results of the answer or the question and it showed which one would be better and it asked which one is better and why is it better and stuff and it already had the results so I could say which one is better. It was an easier question. (Post interview)

A similar positive response was given by Chloe (NES) that question 4 (modified version) was clearer than the original version. She explains:

You can actually see what's going on and it tells you like how many days it was and everything. It tells you everything you need to know to answer the questions. (Post interview)

In contrast, the original version of question 4 (see Appendix 1, Version B) posed more challenges for Joaquin (ELL). One reason for depicting the challenge from ELLs' viewpoint in this study is helping teachers understand their learners more fully. While attempting to answer question 4 (original version) Joaquin stated, "I don't get this question, what is this answer means, because, this question means" (Interviewer 2). Furthermore, Joaquin was interpreting question 4 incorrectly based on what the question meant. The information provided in the question stated two different groups A and B including males and females of different ages, but Joaquin said that data were different for old and young people, which was not included anywhere in the question. Joaquin said, "They like old people and young people . . . like um like they have different problems. . . . They like they have the cough but one's like really bad and one is not really bad . . . that's the similar groups." When the interviewer asked the reasoning for one group receiving medicine with cough syrup and not the other, Joaquin explained, "because Group A is like really bad and these is not really bad so they don't have to um drink or eat the same medicine as them." These excerpts suggested that Joaquin was unable to comprehend the information as given and misinterpreted the science idea (purpose of a control group) in the question.

Although the modified versions were easier to comprehend for both ELLs and NESs, ELLs struggled to pronounce certain words in both versions. This was evident from the conversation between interviewer 1 and Fernando (ELL) while he was attempting to answer question 4 (modified version).

Fernando: A drug company, how do you say that?

I: Developed.

Fernando: Developed (pronounced a little off, voice seemed lower and word dragged out) a new medicine to cure coughs (?) (Voice went up in pitch to make statement into a question.)

Engaging with complex written assessments was difficult and frustrating for the ELLs in this study. The original assessment tasks were the most challenging, but the modified assessments still posed difficulties. While reading the assessments aloud, the challenge was seen in the way ELLs verbalized words, hesitated, repeated, and did not understand particular words and ideas. For example, Joaquin on original question 4 stated:

Joaquin: Scientists have performed a trial of a new cough medicine. They decide a group they divide a group of [partides]?

I: Patients.

...

Joaquin: Plain syrup that did not contain any cough medicine. Everyday for 4 days the scientists int . . . intervi

I: Interviewed.

Joaquin: Interviewed the paint (.5 s pause) the um patients to find out whether their coughs were were as frequent and as serious. They also asked the patients if they had any new health problems while taking the medicine. The following ta . . . table summarizes the data.

...

Joaquin: (2 s pause) They (inaudible) with medicine. . . . They (inaudible) like old people and young people like (.5 s pause) um (1 s pause) like they have different problems.

The modified versions provided prompts that assisted the ELLs and NESs. For example, Fernando reads the question, "[scientist perform it] a brand of this medicine to test how effect the new medicine. It's in correct coughs. So this talks about cough medicine, about cough?" The researcher replied yes and asked what it says about the cough medicine. The student paused for 16 s, leading the researcher to prompt:

I: Like can you read these pictures?

Fernando: Group A received [syrup] with cough medicine. Group B received syrup with, without cough medicine.

Fernando showed frustration throughout the session, but in this instance the prompt provided a way forward, and he was able to describe what was shown in the labeled diagram with less effort.

I: (5 s pause) ok (5 s pause)

Fernando: So why do we do, what do we do?

I: So um, keep reading.

Fernando: At the end of 1 week the [scientist] ask the p—I—ants

I: Patients.

Fernando: Is your cough better, the same, or worse? Do you have any side effects should (ask the scientist or) (starts to mumble and speed slows near end of sentence).

The other students showed the same patterns of difficulty, yet more facility with the modified versions. Table 3 shows the difficulty associated with student responses for question 4 from both modified and original assessments.

Table 3. Difficulty associated with both modified and original assessments (question 4)

Difficulty	Excerpts
Original version	
(1) The student struggled to pronounce patient, received, plain, interviewed, advantages	<p>Joaquin: Scientists have performed a trial of a new cough medicine. They decide a group they divide a group of [partides]</p> <p>Joaquin: Patients with a bad cough into two similar groups. Each group includes males and females and people of difference age. Group A rec . . recide</p> <p>Joaquin: received cough syrup. Group B received plan . . planing . . plan</p> <p>Joaquin: Plain syrup that did not contain any cough medicine. Every-day for 4 days the scientists int . . intervi</p> <p>Joaquin: (Read through that) Should the cough medicine be sold? Be sure to include the adv . . advan</p>
(2) The student did not know the meaning of specific words, for instance "trial"	<p>Interviewer: ok. So let's go back and look at this sentence by sentence. This first sentence: Scientists have performed a trial of a new cough medicine. What does that mean?</p> <p>Joaquin: Um . . (1) that um they made a new medicine.</p> <p>Joaquin: Okay.</p> <p>Joaquin: to make them feel better</p> <p>Interviewer: Alright. What's a trial? What do you think that means?</p> <p>Joaquin: (2) They [] with medicine.</p>

Table 3 continued next page

Table 3. Continued

Difficulty	Excerpts
(3) The student struggled to comprehend the question. The student interpreted the question incorrectly.	<p>Interviewer: ok. They do. They divided a group of patients with a bad cough into two similar groups.</p> <p>Joaquin: they () like old people and young people like (.5) um (1) like they have different problems</p> <p><i>(The data revealed that the student interpreted incorrectly because the patients in both groups had the same problem of cough.)</i></p> <p>Interviewer: Mmhm.</p> <p>Student C1: They like they have the cough but one's like really bad and one is not really bad.</p> <p>Interviewer: Okay.</p> <p>Joaquin: (so that's the similarity)</p> <p>Interviewer: that's the similar groups?</p> <p>Joaquin: mhm</p> <p>Interviewer: Why do you suppose they did that?</p> <p>Joaquin: um because Group A is like really bad and these is not really bad so they don't have to um drink or eat the same medicine as them.</p> <p>Interviewer: okay (1) um. They also asked the patients if they had any new health problems while taking the medicine.</p> <p>Joaquin: Mm (2) I don't get that one</p>
Modified version	
(1) The student could not pronounce the words: Developed, Patients, Tradeoff	<p>Fernando: A drug company, how do you say that?</p> <p>I: Developed</p> <p>Fernando: Developed ((pronounced a little off, voice seemed lower and word dragged out)) a new medicine to cure coughs (?) (voice went up in pitch to make statement into a question)</p> <p>Fernando: At the end of 1 week the <scientist> ask the patients (pronounced more "p-I-ants")</p> <p>Fernando: Because, I don't know. (8s pause) What's (tree house)?</p> <p>I: Um, to include the good reasons and the bad reasons. Tradeoffs</p>

The post-interview responses also revealed that the students found the modified versions more accessible than the original questions with more text. Antonia (ELL) stated that the original version of question 6 (see Appendix 1, Version A) appeared difficult to answer, "Because they're asking me like words that sometimes I don't understand and I don't know how to pronounce it." Moreover, she stated that she had not seen a zebra mussel before and that added difficulty to the question. Dave (NES) agreed. Findings suggested that modifications made the questions more accessible for ELLs and NESs alike. Overall, we found that modifications supported and did not impede both groups of students.

Research Question 2: How does Scaffolding in the Modified Assessment Tasks Support ELLs (and NESs)?

We found three ways that the scaffolding of the modified assessment tasks assisted all students to: (1) increase comprehensibility of the questions, (2) help elicit responses, and

(3) assist in visualizing and organizing thoughts. Each category is presented separately in the sections that follow.

Increasing Comprehensibility

The data analyses of the written assessments, think aloud protocols, and post-assessment interviews showed that the students had more difficulty in understanding the original questions than the modified ones. The modifications that provided structure by division into subparts, graphic organizers, and reduction of abstraction supported both ELLs and NESs by increasing the comprehensibility of the questions. For example, modified question 2 with subsections A, B, and C was easier for students to interpret. We observed that students could focus on each part individually and were more successful than when reading the entire question at one time. We similarly saw Joaquin struggle with question 4 (original version) and incorrectly interpret the difference between Group A and B as old and young people. Because he struggled to make connections between sentences, the interviewer read the question for him. Even after a lot of support, Joaquin faced challenges to understand the demands of the original question, as illustrated below.

I: All right. What's a trial? What do you think that means?

Joaquin: (2 s pause) They [] with medicine.

I: OK. They do. They divided a group of patients with a bad cough into two similar groups

Joaquin: they like old people and young people like (.5 s pause) um (1 s pause) like they have different problems

...

Joaquin: That . . uh . . Group A got cough syrup but it had medicine inside it

I: OK

Joaquin: and group B did receive plain syrup but it didn't have medicine in it.

I: Why do you suppose they did that?

Joaquin: um because Group A is like really bad and these is not really bad so they don't have to um drink or eat the same medicine as them.

I: They also asked the patients if they had any new health problems while taking the medicine.

Joaquin: I don't get that one.

In contrast, the modified version of question 4 seemed to aid comprehension because of division into distinct parts. Fernando faced less difficulty in answering this version:

I: So what do you think would happen, like, if you have these two types of syrups? Can you explain to me what you think this question is asking?

Fernando: It's asking about, how do you FEEL, if, if you feel like this d-d-dizziness or stomach, or, or cough, they give you like, something for you so you can feel better

I: OK, so which medicine would you want to take if you had a cough and wanted to feel better?

Fernando: A

I: Why did you say A?

Fernando: Group A receive syrup with cough medicine

NESs also expressed challenges with the original versions of the questions. Owen (NES) stated that question 6 (original version) was the most difficult one for him to answer. When asked why, he replied, "I think it is because the question was kind of long and you have to read quite a bit. I think it might be easier to answer if the questions were spread apart a little more" (Post interview). A similar response was given by Antonia (ELL) about question 2 (original version).

We also found that the modified versions facilitated comprehension for both NESs and ELLs. This is evident in the conversation between the researcher and Fernando while answering question 2 (modified version) about stopping an antibiotic treatment:

Fernando: She can, probably get more sick.

I: OK.

Fernando: So, Rida stop taking the antibiotics. Decide yes or no and explain your decision. Be sure to include any of . . .

I: So what's that asking you?

Fernando: It's like the same question.

I: OK, so, should Rita stop taking the antibiotics?

Fernando: No

I: OK, and why not?

Fernando: Because it's a good thing if she take it because it's what the doctor tell her and (that's good) she can take it.

Likewise, NESs were more comfortable with the modified version of question 2. Dave was able to understand the question and gave his own example as part of the reasoning: "When I take medicine I don't like it, but I know it will make me better so I still have to take it." Question 2 included modifications that removed abstract language at the beginning of the question and instead began with a concrete situation of Rita feeling sick. It also provided structure through multiple prompts and the addition of a graphic organizer. These changes appeared to help all students reduce frustration and better understand the task. Thus, we found that the modified versions of the assessments were more comprehensible for both ELLs and NESs.

Eliciting Responses

Scaffolding of modified items supported both ELLs and NESs to formulate answers to assessment prompts. One form of scaffolding we found effective was division of the question into multiple prompts. For example, modified question 2 seemed to help students to focus on one reasoning task per prompt. The data revealed that the ELLs did not answer all the parts of original question 2 and that they missed the last part of the question. In contrast, they answered all parts of modified question 2 that provided multiple prompts as well as a graphic organizer. Fernando (ELL), for instance, was able to answer all parts of question

2 (modified) easily. He also reported that the modified version was less intimidating. Likewise, Dave (NES) found division of the question into subparts helpful. He stated, "I liked it split up because you wouldn't have to go back and look at the other part from the other question."

The structural supports helped both ELLs and NESs to focus on the underlying meaning in the question and attend to important aspects to respond appropriately. On question 2 of the modified version, Owen interpreted the question in terms of intended outcomes stating, "Rita is taking antibiotics for a sickness that she has. But it's making her sick. And if she stops taking the medicine, she will continue being sick with what she had before." Similar results were found with Fernando (ELL) while answering this question. He stated, "When I take medicine I don't like it, but I know it will make me better so I still have to take it." Fernando demonstrated he knew that antibiotics treat illness but can also hurt one's stomach. Overall, the structure of multiple prompts and graphic organizers in the modified questions helped both ELLs and NESs to elicit more meaningful responses to the assessment tasks.

Visualization and Organization

Scaffolding of modified items supported both ELLs and NESs in terms of visualization and organization of their thoughts. We found that the visual tools in the scaffolding, such as a diagrams, T-charts, pictures, or bulleted lists, assisted students in reducing their reading time and organizing their ideas. Antonia (ELL) said that question 3 was easiest because of the T-chart, which was rated as the most helpful of all supports within the assessments by the students. T-charts appear to help students particularly with reasoning about a dichotomy (and would not be relevant for any assessment). According to the analysis, Antonia was able to write in the provided box while considering the question, which aided her in organizing her responses. Also, she stated, "When you see chart, it helps you to say what the good things about it are. It is more organized. Helps me write answer better . . ." Also, NESs did not have problems in working with the diagram (question 1, modified version). In addition, during interviews students stated that additional scaffolding tools would help clarify the assessments. Students said, for instance, the original version of question 6 looked complex because of a lot of reading. Bulleted lists might have made this question clearer. Thus, scaffolding through visual tools helped both NESs and ELLs.

Photographs were another type of visual support that students found helpful. Owen stated that question 6 (original version) was the most difficult for him to answer: "The question was kind of long and you have to read quite a bit." However, introducing photographs in question 6 (modified version) was very helpful for both ELLs and NESs. The students stated that photographs provided in the modified version allowed them to visualize the content of the question. Antonia (ELL) mentioned that she liked pictures because it helped her to imagine the zebra mussels, an organism with which she was unfamiliar. On the other hand, the post-interview with Fernando (ELL) revealed that question 6 (original version, without pictures) was the hardest for him. The analysis of the students' responses indicated that the complex language of the question caused the students to struggle to understand. The interviewer had to explain the meaning of terms to the students. When questioned by the interviewer about the difficulty level of original question

6, several students mentioned that they had not seen zebra mussels before and thus needed to see pictures. Overall, scaffolding through visual tools in the modified assessment tasks helped students visualize the scientific context of the questions and organize their thinking, in comparison to the original assessment tasks.

Implications for Teachers

This glimpse into how ELLs tackle language-rich science assessments with and without scaffolding holds important implications for teachers, teacher educators, and professional developers. We discuss these in terms of teachers' views of students, views of assessment, assessment strategies, and assessment resources. The methods in the study provide a model for both sensitizing teachers and building practical skills for modifying assessments equitably for ELLs.

How teachers view ELLs is vitally important to how their learning needs are addressed in the classroom. Most classroom teachers, who are predominantly White, report that they are not well equipped to teach ELLs or attend to their specific needs (Lewis-Moreno, 2007; NCES, 2010). Teachers might view language minorities as having deficits rather than recognizing the wealth of effort second language learners must bring to learning in a content area. This study depicted specific examples of ways ELLs struggle with written science assessments, while also showing ways a variety of scaffolds assist ELLs. It is worth noting the results of the first research question showed that ELLs struggled with the assessments and felt challenged and frustrated. Developing a more sensitive nature to "the immense challenge of being an ELL" and understanding learning and assessment needs is a major undertaking for teachers and teacher developers (Moore, 2007). By providing detailed examples of ways ELLs are challenged by the language in assessments, such as those in this study, teacher developers can help teachers alter their sensitivity to ELLs' needs. In a recent study, the first author incorporated equitable instruction and assessment into a methods course for preservice teachers. Findings showed that preservice teachers gained a new view of the difficulties ELLs face, their abilities, and equitable assessment strategies (Siegel, 2014).

A second implication of the current study relates to teachers' views of assessment and grading. What are the implications of the students receiving failing marks on all the original assessments? Modifying assessments is a challenging task, and few teachers use assessments that are appropriately modified for ELLs. As this small study and others have shown (Abedi et al., 2004; Siegel, 2007), modified assessments assist both ELLs and NESs. Modifications to increase equity usually help all students and are an important aspect of standardized test research for this reason (Butler & Stevens, 2001). Thus, we want to encourage teachers and those who work with teachers to learn about equitable ways to modify assessments for students, use them, and support language development in contrast to criticizing or ignoring it. It is fair to focus on content for ELLs while also correcting examples of language, being explicit about the reason, and asking them to correct the rest. As Lee, Quinn, and Valdes (2013, p. 9) recommends, "Participation in science and engineering practices should be expected of all students, and ELLs' contributions should be accepted

and acknowledged for their value within the science discourse, rather than critiqued for their 'flawed' use of language."

A third implication involves assessment strategies and power. As Moore (2007) points out, teachers need to see the role of language and discourse as a gatekeeper, as power. Teachers can explicitly teach the rules of power to students (Moore, 2007). For example, what are the rules of succeeding on written classroom assessments that address the NGSS? From this study, teachers can help students learn how to interpret written assessments and how to benefit from the scaffolds provided. The techniques for interpreting diagrams and pictures, finding main points, and seeing clues in questions will help all students be more effective learners. It is helpful to show students the implicit ideas the teacher expects on written responses. Teachers should discuss various samples of student work and how some address the questions more clearly and persuasively. This would help students recognize the hidden rules of assessment in their classes and become more self-regulated learners. Furthermore, the design of assessment, as this study demonstrated, is a fundamental challenge to ELLs' learning and teachers' assessment practice. Discussing how assessments are designed and what the teacher intended is not common practice in classrooms but could have a profound effect on learners.

Fourth, while equitable assessment resources are scant, it is critical for teachers to know about and gain access to what is available. This study provides assessment tasks addressing the NGSS (see Appendix 1) and effective modifications to these assessments (see Appendix 2 for a list of all modifications by question number). In summary, specific types of modifications we found useful in this study included:

- Visual tools, such as photographs, diagrams, and bulleted lists
- Structural scaffolds, such as multiple prompts, and graphic organizers
- Text modifications, such as reduction of words and matching terms to the language of instruction.

Scaffolds should shape learners' engagement with assessments. Findings in this study indicate specific ways scaffolds transform science assessment tasks. The ability of scaffolding to change the way a student envisions a task and engages in reasoning was evident. We found effects on: (1) the input, or comprehension side; (2) the output, or production side; and (3) the thinking that is perhaps in between, in terms of visualizing the context of the question, organizing ideas, and monitoring solutions. We found that structural modifications aided in comprehension and elicitation, and visual tools aided in organization of ideas. The study included multiple modifications per task, and thus there could be effects from other modifications that we could not detect. Investigating in more detail how specific types of scaffolding assist students on written assessment tasks will be a useful next step for researchers.

Moreover, the original and modified assessment tasks that align with the NGSS from this study can be employed to study assessment design and use in middle school classrooms. Teachers could use the methods of data collection and analysis from this study to conduct action research to hone their skills in assessment modification for ELLs. Having access to both original and modified tasks can serve as a resource to provide scaffolding

for certain students, or to fade scaffolding over time. Also, by examining data through the lens of the equitable assessment framework, evidence was amassed showing that the principles are useful for both ELLs and NESs, suggesting that the framework would be useful for developing new assessment tasks. Further assessment resources that attend to the needs of ELLs can be found online or by contacting the authors referenced (Buxton et al., 2013; Siegel et al., 2005).

Conclusion

Limited studies focus on the language opportunities and challenges offered by science assessment tasks for ELLs (see, e.g., Shaw, 1997; Shaw et al., 2010; Siegel, 2007). While teachers use scaffolding to help ELLs learn, research has not examined how scaffolding might support or impede learners specifically on written classroom science assessment questions. Standardized test questions have received more attention, but we focused on complex formative tasks that a science teacher would appreciate using in a classroom. The main purpose of the study was to take a first step in gaining insight to how scaffolding is useful through the voices of students. This study demonstrated that structural modifications, such as multiple prompts and graphic organizers aided in comprehension. Structural scaffolding also enhanced students' abilities to respond meaningfully to assessments. Visual tools, such as photographs and diagrams, were helpful in visualizing and organizing ideas. Overall, we found that scaffolding on the modified assessment tasks aided ELLs and NESs to comprehend, visualize, and respond to challenging life science assessments.

Acknowledgments – We would like to thank Jerome Shaw for comments on an earlier version of this manuscript and Jenna Krueger and Emily Lloyd for transcription and research assistance through the College of Education honors program. Also, we are grateful to the students and teachers of the participating schools. Partial funding was provided by MU Research Council and Sandra K. Abell's MU Science Education Center.

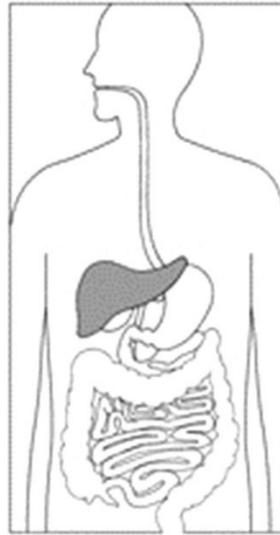
Appendix 1

Version A

Question 1

NGSS: MS-LS1-3; DCI: LS1.A; CC: Structure and function

Imagine taking a bite of food. Follow the food through the digestive system, explaining where the food goes, what organs are involved, and what is taking place as it moves along.



Question 2

NGSS: MS-LS1-3; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions; CC: Cause and Effect, Structure and Function

Rita felt sick and went to the clinic. Her doctor said she had a bacterial infection and gave Rita a prescription for antibiotics. She was told to take the antibiotics for 10 days. Rita took the antibiotics for 3 days and now feels completely better. The antibiotics upset Rita's stomach, so she wants to stop taking them.

- A. What are some good and bad things about stopping the full course of antibiotics? What are some good and bad things about continuing to take the full course of antibiotics? Write your answers in the table:

Good things		Bad things	
Stop taking antibiotics	Continue taking antibiotics	Stop taking antibiotics	Continue taking antibiotics

- B. If Rita stops taking the antibiotics, what effect could it have on the bacteria causing Rita’s infection? *Be sure to discuss the scientific principles that explain this effect.*
- C. Should Rita stop taking the antibiotics? Decide yes or no, and explain your decision. Be sure to include any trade-offs involved.

Question 3

NGSS: MS-LS1-3; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions; CC: Cause and Effect, Science Is a Human Endeavor

Joe has a cough that may be infectious. Should people with infectious diseases be quarantined from other people? Use the T-chart to organize your thoughts.

Should be quarantined	Should not be quarantined

Should Joe be quarantined?

Question 4

NGSS: MS-LS1-3; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions, Planning and Carrying out Investigations; CC: Cause and Effect, Science Is a Human Endeavor

A drug company developed a new medicine to cure coughs. Scientists performed a trial of this medicine to test how effective the new medicine is in curing coughs.



Group A received syrup with cough medicine.



Group B received syrup **without** cough medicine.

At the end of 1 week, the scientists asked the patient Fernando

- Is your cough better, the same, or worse?
- Do you have any side effects, such as dizziness or upset stomach?

Table 4 gives the results.

Table 4. New cough medicine results (number of people)

Group	All patients	Cough is better	Cough is the same	Cough is worse
A (syrup with cough medicine)	50	40	5	5
B (syrup without cough medicine)	50	20	25	5

- Why did group B get syrup without cough medicine?
- Compare the two groups. How well does the new medicine work in curing coughs? Be sure to use data from Table 4 in your answer.

Table 5. New cough medicine—more results (number of people)

Group	Side effects (dizziness and stomach upsets)
A (syrup with cough medicine)	10
B (syrup without cough medicine)	2

- What are some good and bad things about the new medicine? Write these in Table 6. Be sure to use results from Tables 4 and 5 in your answer.

Table 6. Good and bad things about the new cough medicine

Good things	Bad things

- D. Even though there are side effects, would you buy this medicine? Explain why and be sure to include the trade-offs.

Question 5

NGSS: DCI: LS1.B; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions, Planning and Carrying out Investigations; CC: Cause and Effect, Science Is a Human Endeavor; Common Core: Write Arguments Focused on Discipline Content

Joe's mom had a serious heart problem that doctors now think may have been caused by a genetic disease.

- A. Can you catch genetic diseases from other people the same way you can catch the flu? Explain.
- B. Should Joe be tested for the genetic disease his mother may have?

Give reasons for testing and against testing:

Reasons to be tested	Reasons not to be tested

Should Joe be tested? _____

- C. Pretend you are Joe's friend. Write a letter to Joe telling him whether you think he should be tested. Be sure to include both sides of the issue.

Question 6

NGSS: MS-LS2-1, MS-LS2-2, MS-LS2-4, MS-LS2-5; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions; CC: Cause and Effect, Science Is a Human Endeavor

Zebra mussels are an introduced species that have invaded US lakes and rivers. These mussels normally live in Russia. These small mussels can grow on almost any surface. Scientists believe that some of these mussels attached to the bottom of boats and anchors bringing them to Missouri. Zebra mussels cause problems because they out-compete our native mussels for food. They are so sharp that they can cause beaches to be too dangerous to swim in without shoes. These mussels can also grow so closely together that they often

block off pipelines, preventing cities from accessing water and limiting hydroelectric power generation.

- A. Describe the impact of zebra mussels on Missouri ecosystems in terms of competition and the mussels' effects on manmade structures.
- B. How do you think these mussels may affect fisheries in the area?
- C. Considering what you know about how zebra mussels can spread, how would you plan to limit the spread of zebra mussels to new lakes and rivers? What steps might you suggest scientists take to begin permanently removing these mussels from the ecosystem?

Version B

Question 1

NGSS: MS-LS1-3; DCI: LS1.A; CC: Structure and function

Imagine that you have just taken a bite of lunch. Label a diagram or chart of the path that the food takes as it moves through the digestive system, using the words provided. Briefly tell what happens to the food as it passes through each organ.

Word bank:

mouth	liver	kidney
teeth	large intestine	rectum
saliva	small intestine	
esophagus	pancreas	
stomach	gall bladder	

Question 2

NGSS: MS-LS1-3; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions; CC: Cause and Effect, Structure and Function

Rita began taking a 10-day treatment of antibiotics 3 days ago. The antibiotics worked quickly, and Rita feels completely better after only 3 days. Antibiotics upset Rita's stomach, so she wants to stop taking them.

Should Rita stop taking the antibiotics or finish the treatment? Explain the advantages and disadvantages of stopping and of continuing the antibiotics.
Be sure to include your final recommendation, any trade-offs involved, and your reasons for your decision.

Question 3

NGSS: MS-LS1-3; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions; CC: Cause and Effect, Science Is a Human Endeavor

Joe has a cough that may be infectious. Should people with infectious diseases be quarantined from other people. Use the T-chart to organize your thoughts.

Should be quarantined	Should not be quarantined

Should Joe be quarantined?

Question 4

NGSS: MS-LS1-3; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions, Planning and Carrying out Investigations; CC: Cause and Effect, Science Is a Human Endeavor

Scientists have performed a trial of a new cough medicine. They divided a group of patients with a bad cough into two similar groups. Each group included males and females and people of different ages. Group A received cough syrup. Group B received plain syrup that did not contain any cough medicine. Every day for 4 days, the scientists interviewed the patients to find out whether their coughs were as frequent and as serious. They also asked the patients if they had any new health problems while taking the medicine. The following table summarizes the data.

Group	Total number of patients	Number who feel better	Number who feel the same	Number who feel worse	Number with side effects (dizziness and upset stomach)
A (cough syrup)	50	40	5	5	10
B (plain syrup)	50	20	25	5	2

- A. Analyze the data to form a conclusion about how well the medicine works.
- B. Should the cough medicine be sold? Be sure to include the advantages and disadvantages of **BOTH** choosing to sell and choosing not to sell the medicine. Explain the trade-offs of your final decision.

Question 5

NGSS: DCI: LS1.B; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions, Planning and Carrying out Investigations; CC: Cause and Effect, Science Is a Human Endeavor; Common Core: Write Arguments Focused on Discipline Content

Joe's mom had a serious heart problem that doctors now think may have been caused by a genetic disease.

- A. Can you catch genetic diseases from other people the same way you can catch the flu? Explain.

Answer: _____

Explain:

What do you know about how people get genetic disease?	What do you know about how people get nongenetic diseases?

- B. Should Joe be tested for the disease his mother may have? Give reasons for testing and against testing and then make a decision.

Should Joe be tested? _____

Reasons

- C. Pretend you are Joe's friend. Write a letter to Joe telling him whether you think he should be tested. Be sure to include both sides of the issue.

Dear Joe,

I think that you [insert opinion] be tested for a genetic disease. I believe this because [insert reasons that support your opinion]. Although there are other reasons to consider, like [insert reasons that are against your opinion]. I still feel that you [insert opinion].

Sincerely,

[Insert your name]

Question 6

NGSS: MS-LS2-1, MS-LS2-2, MS-LS2-4, MS-LS2-5; SEP: Engaging in Argument from Evidence, Constructing Explanations and Designing Solutions; CC: Cause and Effect, Science Is a Human Endeavor

Zebra mussels are an introduced species that have invaded US lakes and rivers. These mussels normally live in Russia. These small mussels can grow on almost any surface. Scientists believe that some of these mussels attached to the bottom of boats and anchors bringing them to Missouri. Zebra mussels cause problems because they out compete our native mussels for food. They are so sharp that they can cause beaches to be too dangerous to swim in without shoes. These mussels can also grow so closely together that they often block off pipelines preventing cities from accessing water and limiting hydroelectric power generation.

- A. What has been the impact of the zebra mussel on ecosystems in Missouri in terms of the following:
 - Competition with other mussel species
 - Effects on manmade structures and equipment (dams, boats, etc.)
- B. How do you think these mussels may affect fisheries in the area?
- C. Considering what you know about how zebra mussels can spread, how would you plan to limit the spread of zebra mussel to new lakes and rivers? What steps might you suggest scientists take to begin permanently removing these mussels from the ecosystem?

Appendix 2

Description of modifications by question			
Question	Version A	Version B	Modification
1	Modified		Reduction of words; addition of visual support
2	Modified		Matching the language of instruction more precisely; linguistic simplification of vocabulary and syntax; addition of graphic organizer; breaking of question into parts
3	Same	Same	Graphic organizer
4	Modified		Reductions of words in the item stem; addition of visual supports; matching the language within the item more precisely; reduction of nonessential information; replacing sentences with bulleted lists; linguistic simplification of vocabulary and syntax; use of bold type for emphasis; addition of graphic organizer
5		Modified	Addition of graphic organizer; linguistic simplification; detailed guide/framework to support answering; matching the language within the item more precisely
6		Modified	Addition of visual support; brief note describing each visual support; reduction of words in the item stem; replacing a long question with two simple questions

References

- Abedi, J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33, 4–14.
- Abedi, J. (2009). *Guidelines for assessing English learners in academic content areas: Linguistic modifications of assessment*. Paper presented at the CCSSO National Conference on Student Assessment, Los Angeles, CA.
- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74, 1–28.
- Abedi, J., Lord, C., Hofstetter, C. H., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16–26.
- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205–221). London: Springer.
- Baxter, G. P., Shavelson, R. J., Goldman, S. R., & Pine, J. (1992). Evaluation of procedure based scoring for hands-on science assessment. *Journal of Educational Measurement*, 29, 1–17.
- Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20, 463–494.
- Bryan, L. A., & Atwater, M. M. (2002). Teacher beliefs and cultural models: A challenge for science teacher preparation programs. *Science Education*, 86, 821–839.
- Butler, F., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K–12: Current trends and old dilemmas. *Language Testing*, 18, 409–427.
- Buxton, C. A., Allexsaht-Snider, M., Suriel, R., Kayumova, S., Choi, Y.-J., Bouton, B., & Baker, M. (2013). Using educative assessments to support science teaching for middle school English language learners. *Journal of Science Teacher Education*, 24, 347–366.
- CRESST. (2001). *Policy brief no. 4*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Hazel, E., Logan, P., & Gallagher, P. (1997). Equitable assessment of students in physics: Importance of gender and language background. *International Journal of Science Education*, 19, 381–392.
- Hobsbaum, A., Peters, S., & Sylva, K. (1996). Scaffolding in reading recovery. *Oxford Review of Education*, 22, 17–35.
- Jenelle, R. (2004). Like everybody else: Equalizing educational opportunity for English Language Learners. *TESOL Quarterly*, 38, 43–66.
- LaCelle-Peterson, M. W., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review*, 64, 55–75.
- Lee, O. (2001). Culture and language in science education: What do we know and what do we need to know? *Journal of Research in Science Teaching*, 38, 499–501.
- Lee, O. (2005). Science education and English language learners: Synthesis and research agenda. *Review of Educational Research*, 4, 491–530.
- Lee, O., Luykx, A., Buxton, C., & Shaver, A. (2007). The challenge of altering elementary school teachers' beliefs and practices regarding linguistic and cultural diversity in science instruction. *Journal of Research in Science Teaching*, 44, 1269–1291.
- Lee, O., Quinn, H., & Valdes, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for common core state standards for English language arts and mathematics. *Educational Researcher*, 42, 223–233.

- Lewis-Moreno, B. (2007). Shared responsibility: Achieving success with English-language learners. *Phi Delta Kappan*, 88, 772–775.
- Lincoln, Y. S., & Guba, E. G. (1985). Establishing trustworthiness. In Y. S. Lincoln & E. G. Guba (Eds.), *Naturalistic inquiry* (pp. 289–331). Newbury Park, CA: Sage.
- Lyon, E. (2013). Assessment as discourse: A pre-service physics teacher's evolving capacity to support an equitable pedagogy. *Education Sciences*, 3, 279–299.
- Moore, F. M. (2007). Language in science education as a gatekeeper to learning, teaching, and professional development. *Journal of Science Teacher Education*, 18, 319–343.
- National Center for Education Statistics (NCES). (2010). *The condition of education 2010*. Washington, DC: IES.
- National Research Council (NRC) (Achieve, NSTA). (2013). *Next generation science standards*. Retrieved from <http://www.nextgenscience.org/next-generation-science-standards>
- Palincsar, A. S. (1998). Keeping the metaphor of scaffolding fresh: A response to C. Addison Stone's – The metaphor of scaffolding: Its utility for the field of learning disabilities. *Journal of Learning Disabilities*, 31, 370–373.
- Shaw, J. M. (1997). Threats to the validity of science performance assessments for English language learners. *Journal of Research in Science Teaching*, 34, 721–743.
- Shaw, J. M., Bunch, G. C., & Geaney, E. R. (2010). Analyzing language demands facing English learners on science performance assessments: The SALD framework. *Journal of Research in Science Teaching*, 47, 909–928.
- Siegel, M. A. (2007). Striving for equitable classroom assessments for linguistic minorities: Strategies for and effects of revising life science items. *Journal of Research in Science Teaching*, 44, 864–881.
- Siegel, M. A. (2014). Developing preservice science teachers' expertise in equitable assessment. *Journal of Science Teacher Education*, 25, 289–308.
- Siegel, M. A., Markey, D., & Swann, S. (2005). *Life science assessments for English learners: A teacher's resource*. Berkeley, CA: University of California. <http://www.lmri.ucsb.edu>
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, 22, 371–391.
- Siegel, M. A., Wissehr, C., & Halverson, K. L. (2008). Sounds like success: A framework for equitable assessment. *The Science Teacher*, 75(3), 43–46.
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37, 189–199.
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38, 553–573.
- Vygotsky, L. S. (1978). *Mind in society: The development to higher psychological processes*. London: Cambridge University Press.
- Walqui, A. (2003). *Conceptual framework: Scaffolding instruction for English learners*. San Francisco, CA: WestEd.
- Wiley, T. G., & Wright, W. E. (2004). Against the undertow: Language-minority policy and politics in the "age of accountability." *Educational Policy*, 18, 142–168.
- Wong-Fillmore, L. (2007). English learners and mathematics learning: Language issues to consider. *Assessing Mathematical Proficiency*, 53, 333–344.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89–100.