

Article

Developing a Technology-Based Classroom Assessment of Academic Reading Skills for English Language Learners and Teachers: Validity Evidence for Formative Use

Mikyung Kim Wolf *  and Alexis A. Lopez 

Educational Testing Service (ETS), Princeton, NJ 08541, USA; alopez@ets.org

* Correspondence: mkwolf@ets.org

Abstract: In U.S. K-12 schools, adequate education of English language learner (EL) students, particularly to support their attainment of English language and literacy skills, has attracted heightened attention. The increased academic rigor as well as sophisticated disciplinary language demands embodied in current academic content standards have posed considerable challenges to EL students. To address students' needs, the present study utilized formative assessment as a means to support the teaching and learning of academic reading skills for EL students. We also endeavored to test our underlying assumption that sound assessment tools would facilitate effective formative assessment processes. In this study, we devised a technology-based assessment tool considering the increasing use of technology in K-12 schools. As a small-scale, exploratory study, we examined the usability and validity of the tool for formative purposes with three ESL teachers and their students (62 EL students) from secondary schools. The results indicated that the tool had the potential to extend teachers' and students' formative assessment practices in principled ways. However, we also found some teachers' misconceptions about the tool's purpose and their limited implementation skills to utilize the tool for formative assessment purposes. Implications for practice and further research are discussed.

Keywords: academic reading skills; classroom assessment; English language learners; formative assessment; technology-based language assessment; validity



Citation: Wolf, Mikyung Kim, and Alexis A. Lopez. 2022. Developing a Technology-Based Classroom Assessment of Academic Reading Skills for English Language Learners and Teachers: Validity Evidence for Formative Use. *Languages* 7: 71. <https://doi.org/10.3390/languages7020071>

Academic Editors: Dina Tsagari, Henrik Böhn, Juana M. Liceras and Raquel Fernández Fuertes

Received: 31 October 2021

Accepted: 17 March 2022

Published: 22 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In kindergarten to Grade 12 (K-12) public schools in the United States, students who are officially designated as an English learner (EL) account for 10% of the total student enrollment (U.S. Department of Education, Office of English Language Acquisition [OELA] 2021). English learners (ELs) are a specific term used in official documents in U.S. K-12 education to refer to students whose home language is not English and whose English language proficiency has not yet developed to meaningfully participate in school settings (Every Student Succeeds Act 2015). Students who are classified as ELs are entitled to receive appropriate instructional services and language support in developing their English language proficiency. U.S. federal legislation explicitly stipulates that states and schools include EL students in their annual statewide assessments in content areas (e.g., English language arts, mathematics, science) as well as in English language proficiency. States are then required to report EL students' academic achievement and progress in English language proficiency attainment as a separate subgroup for accountability purposes. This subgroup reporting unveiled a substantial achievement gap that EL students experience. For example, in the state of California, the EL student group was reported to be about 90 points below meeting the grade-level standard in English language arts compared to the non-EL student group, which was about nine points above the standard in 2019 (<https://www.caschooldashboard.org/>, accessed on 17 September 2021). Further, a large number of EL students were found to be "long-term EL" students, staying in EL designation for over six years (Sugarman and Lee 2017). Providing effective strategies to help EL

students develop appropriate English language proficiency for their academic success has been a pressing issue in U.S. K-12 education.

As a powerful instructional strategy, formative assessment has been gaining renewed interest among researchers and practitioners in contributing to learning outcomes (e.g., [Black and Wiliam 2010](#); [Gan and Leung 2019](#); [Heritage 2013](#); [Ruiz-Primo et al. 2014](#)). By definition, formative assessment involves an ongoing assessment process to identify students' status in relation to learning goals and to provide targeted instruction based on individual students' needs. Considering EL students' diverse backgrounds in terms of their language proficiency levels, formal schooling experience, length of US residence, and cultures, formative assessment can be beneficial for EL students and their teachers to provide tailored instruction to address various EL students' needs.

For effective formative assessment, teachers' assessment knowledge and implementation skills are essential. Previous research has underscored the importance of professional development and support to realize the intent of formative assessment in practice (e.g., [Leung 2004](#); [Tsagari and Vogt 2017](#)). However, teachers' busy schedules add to the challenges of devising or selecting appropriate methods and planning for the execution of effective formative assessment. Furthermore, not all teachers are equipped with the linguistic and language instruction knowledge necessary to implement formative assessment for EL students ([Callahan 2013](#)).

Against this backdrop, in this study, we attempted to devise a classroom-based assessment tool to facilitate effective formative assessment processes for EL students and their teachers. With the development of a prototype assessment tool and its usability study, we aimed to examine the characteristics associated with effective formative assessment in the context of EL education. We anticipated that study findings would provide useful information to enhance such assessment tools and formative assessment processes, thereby contributing to teaching and learning for EL students.

2. A Framework to Develop and Validate an Assessment Tool for Formative Use

In U.S. K-12 education contexts, the concept of formative assessment has been construed differently in practice. Some contend that assessments can be formative assessments insofar as they provide information to plan instruction while others argue that formative assessment is a process rather than instruments ([Bennett 2011](#); [Heritage 2010](#)). In this study, we employed [Popham's \(2008\)](#) definition of formative assessment as "a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics" ([Popham 2008](#), p. 6). As formative assessment is a process, validation of formative assessment must attend to assessment processes that lead to the intended outcomes. That is, fidelity of assessment implementation for formative use is central to validation. For example, evidence related to implementation during formative assessment processes may pertain to how teachers elicit learning evidence, how they interpret the evidence, what feedback they provide for students to progress to the next step toward a learning goal, and how students engage in assessment and feedback. [Shepard \(2009\)](#) stresses this process-focused approach to evaluating the validity of formative assessment. She asserts, "it is the use of an instrument, rather than the instrument itself that must be shown, with evidence, to warrant the claim of formative assessment" ([Shepard 2009](#), p. 33). In the concept of formative assessment focusing on its process, both teachers and students should be actively involved in the assessment process to make formative assessment effective and to result in intended learning outcomes.

Past research on formative assessment elucidated key elements that constitute formative assessment. For instance, [Black and Wiliam \(1998\)](#) delineate that the effective implementation of formative assessment requires the following five elements: (1) setting clear goals; (2) collecting learning evidence through appropriate learning and assessment tasks; (3) communicating assessment criteria; (4) teachers' providing feedback; and (5) students' engaging in self-and peer-assessment. Other researchers also emphasize the importance

of teachers' role in analyzing and interpreting information based on collected evidence and in providing appropriate feedback (Leung and Mohan 2004; Leung and Scott 2009; Rea-Dickins 2001; Ruiz-Primo and Li 2013; Schildkamp et al. 2020). Likewise, students' active role in acting upon feedback while engaging in self-assessment is an important element of formative assessment (Heritage 2013).

Drawing from the previous literature, we adapted a formative assessment framework with three high-level process elements (see Figure 1). This framework was used to develop our assessment tool and initial validation work for the formative use of the tool in this study. As Bennett (2011) maintains, formative assessment is difficult to achieve without sound instrumentation. In an effort to support good instrumentation, we focused on developing an assessment tool that could facilitate the implementation of key formative assessment elements. This tool contained a specific learning goal with the description of a target construct (i.e., academic reading skills), its progression model, assessment tasks, a performance report and resources as a way of feedback, and self-assessment questions. More details of the tool and the targeted construct are described in a later section.

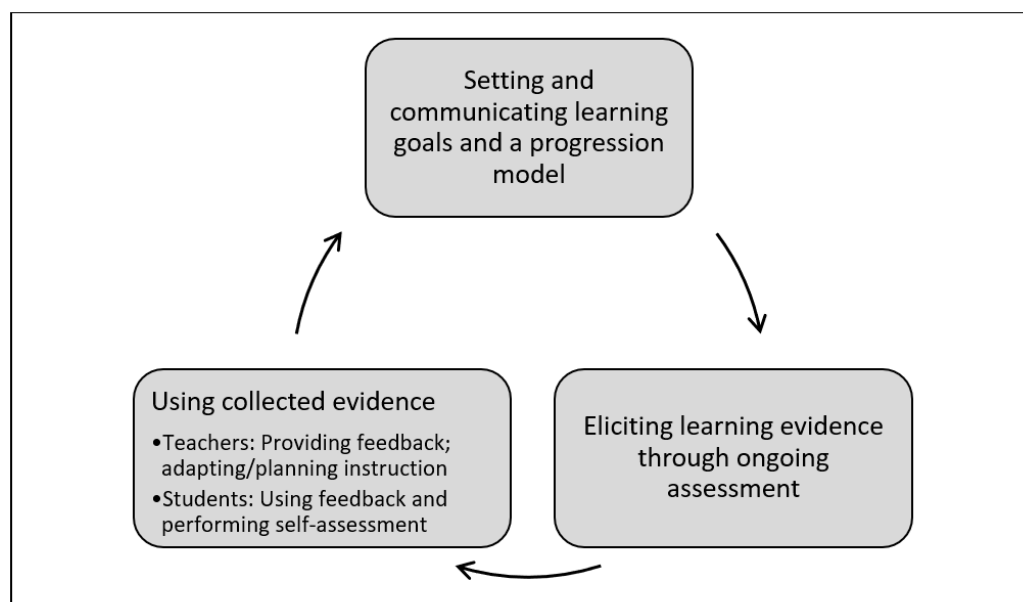


Figure 1. A formative assessment framework to develop and validate the study assessment tool for formative use.

Considering the increased use of technology and computer-based materials in current U.S. K-12 education, we built all of these elements in a computer-based learning management system. Despite a widespread use of technology in schools, technology-based, classroom assessment tools for EL students for formative purposes are scant (Hamill et al. 2019). By developing an assessment tool using technology features (e.g., computer-delivery, multimedia integration, various item/task formats, immediate scoring/feedback), we intended the tool to be easily integrated in current instructional settings. In addition, we aimed to examine any technology-specific advantages or disadvantages that teachers and students might encounter in using such a tool for formative assessment purposes.

3. Academic Reading Skills and Reading Standards in U.S. K-12 School Settings

Since formative assessment should be an integral part of instruction, it is important to align a target construct of assessment with a school's standards and/or curricula in the context of K-12 education. The target construct of academic reading skills in this study was driven by the reading standards adopted widely across U.S. K-12 schools (i.e., Common Core State Standards). In this section, we briefly describe the reading standards on which

this study was based on the academic reading skills expected of students, including EL students, in U.S. K-12 school contexts.

As part of standards-based education reform, U.S. K-12 schools must adopt academic content and English language proficiency standards. These standards then guide what is taught and assessed. In efforts to prepare all students for college and careers, a nationwide initiative took place in U.S. K-12 education a decade ago, resulting in a new set of standards named the Common Core State Standards. Subsequently, many states and schools adopted the Common Core State Standards or revised their existing standards to be similar to the Common Core State Standards. As the name indicates, these standards feature a set of core knowledge and skills for students to achieve in order to be college- and career-ready. In addition to increased academic rigor, the standards expect students to demonstrate sophisticated language use (Bailey and Wolf 2020; Bunch 2013). The ten reading standards of the Common Core State Standards feature analyzing both complex informational and literary texts. They also expect students to be able to integrate and evaluate information and arguments from multiple texts. The ten reading standards are consistent from kindergarten to Grade 12 with a different degree of complexity for each grade level. The reading skills manifested in current standards in U.S. K-12 schools tend to focus on higher-level skills in academic contexts. To support the instruction of EL students, schools' English language proficiency standards were also revised in accordance with the expectations delineated in the Common Core State Standards, emphasizing academic language proficiency to handle materials and tasks in school settings.

Ample research on reading skills has suggested that reading is a multicomponent construct (e.g., Alderson 2000; Koda 2004; Sabatini et al. 2012). Broadly speaking, reading skills involve both lower-level and higher-level skills (Grabe 2009; Saito and Inoi 2017). Lower-level skills include foundational reading skills such as decoding and processing sentence structures (Bernhardt 2011; O'Reilly and Sheehan 2009). Higher-level skills involve building a mental model of comprehended texts including such skills as summarizing, making inferences, and integrating multiple information beyond literal understanding of texts (O'Reilly and Sheehan 2009; Saito and Inoi 2017). While academic reading skills are commonly involved in higher-level skills, lower-level skills are essential to perform higher-level skills. Although the current reading standards in U.S. K-12 schools focus on higher-level reading skills, it is crucial to instruct and assess both lower- and higher-level reading skills in order to support the needs of EL students with a wide range of English language proficiency.

4. The Development of an Academic Reading Assessment Tool for Formative Use

In building a prototype assessment tool, we narrowed the target users as EL students at the intermediate to advanced English language proficiency levels (including long-term ELs) and ESL teachers in secondary school, particularly in Grades 6–8. Reflecting the formative assessment framework (Figure 1), the tool was developed to include four components for teachers and students to engage in: (a) learning goals, (b) activities (assessment tasks), (c) performance results (individual score reports as well as teacher feedback), and (d) next steps and resources (students' self-assessment, planning template, and further practice materials). Figure 2 presents the home page of this assessment tool (i.e., a learning management system) and the features included in each component. This learning management system allows both teachers and students to log on and view all components and features.

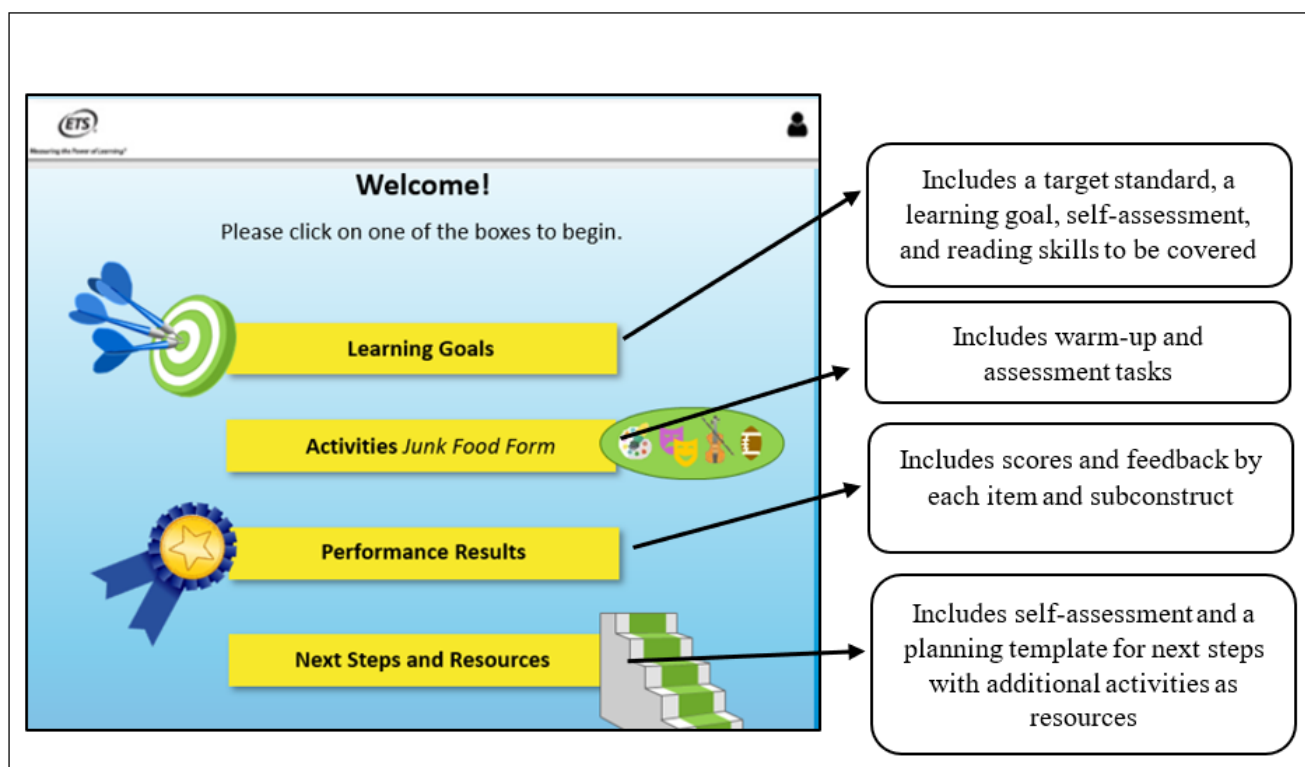


Figure 2. A screenshot of the assessment tool (learning management system) home page and key features.

As described in the previous section, for the assessment tool to be easily integrated with the school curriculum as part of their daily instruction, we selected one specific reading standard as a learning goal. This standard from the Common Core State Standards for English Language Arts-Reading in Grades 6–12 reads, “Delineate and evaluate the argument and specific claims in a text, including the validity of the reasoning as well as the relevance and sufficiency of the evidence”. ([Common Core State Standards Initiative 2010](#), p. 35). The content of this standard was also reflected in some U.S. K-12 English language proficiency standards. The ELPA21 English language proficiency standards that are currently being used in nine states contain the following standard, “Analyze and critique the argument of others orally and in writing” for Grades 6–8 ([Council of Chief State School Officers 2014](#), p. 4). Based on these standards, we defined the learning goal and target construct as being able to comprehend argumentative texts and evaluate authors’ claims along with evidence. We further specified three subconstructs of foundational, literal comprehension, and higher-order reading comprehension skills as a progression model. Then, specific tasks were devised to assess each subconstruct skill. [Table 1](#) summarizes the subconstructs and task types of this assessment tool.

Table 1. The subconstructs and task types of the academic reading assessment tool.

Subconstructs	Task Type Names
Foundational skills: Understanding and using English words (vocabulary) and forms (grammar) in order to interpret the meaning of the sentences	Working with words (3) Working with grammar (4)
Literal comprehension skills: Identifying and understanding an author’s main idea/argument and supporting details/evidence	Understanding main ideas (2) Getting the details (2) Distinguishing facts from opinions (5)
Higher-order comprehension skills: Analyzing an argument structure, making inferences, and making connections between texts in order to evaluate arguments	Working with argument structure (3) Drawing inferences (3) Making connections (3) Evaluating arguments and evidence (2)

Note. The number in the parenthesis indicates the number of items per each task type.

The task name was shown for each item on the computer screen so that both students and teachers were aware of the types of skills in which they were engaged. This design was also intended to display a progression model so that teachers could easily identify students’ current status and consider the next steps in relation to the overall learning goal and targeted construct.


As a way to facilitate students’ understanding of learning goals, students’ self-assessment was constructed in alignment with the subconstruct skills. Self-assessment questions were then embedded in the Learning Goals component shown in Figure 2. Some example of self-assessment questions include *Can you understand difficult words using clues in a text? Can you understand the author’s main opinion or argument? Can you recognize and understand the way that an author organized the text?*

In developing assessment tasks, a few design features were applied to make this tool specific to EL students’ needs. Those features included: (a) providing warm-up tasks in order to activate EL students’ background knowledge on the topic of reading; (b) integrating multiple language skills for students to unpack the passage to build comprehension (e.g., a text-to-speech feature, tasks requiring students to discuss and write about the reading topic); and (c) providing scaffolded tasks while modeling a reading comprehension process (e.g., sequencing the tasks to help students do a close reading of the text). Hence, several warm-up tasks were designed for teachers to select at their discretion. For example, a warm-up task asked students to talk in pairs on a topic relevant to the reading passage prior to reading the passage. The main tasks in the assessment, then, began with introducing two authors and telling students the purpose for reading. Figure 3 displays the screenshots of a few sample tasks in the beginning of the Activities component to illustrate these features. Although it is expected for students in Grades 6–8 to have foundational reading skills, tasks covering foundational reading skills were embedded in the tool as one way of scaffolding for EL students, considering their developing English proficiency. The assessment activities consisted of two parts. Part 1 was designed to be completed collaboratively with a peer and provided opportunities for a teacher to observe and interact with students while eliciting evidence about EL students’ reading skills. Part 2 was an individual assessment based on the same reading passages as in part 1. Appendix A provides some examples of part 1 and part 2 tasks to illustrate reading tasks devised to assess each subconstruct skill.

Warm-up activity sample

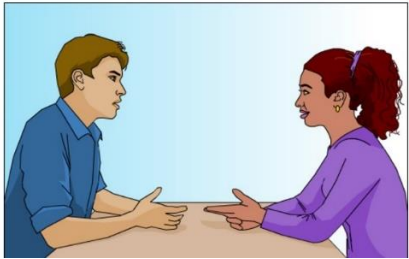

Listening. Learning. Leading.®

Warm-Up #1









Now discuss the following questions with your partner.



- 1) What did you eat for dinner last night?
- 2) Do you think it was a healthy food choice?


Introduction



Listening. Learning. Leading.®

Introduction










Emil Fedor

Hello, I'm Emil Fedor. I wrote an article in Students and Health magazine about why we should not have junk food in schools. I'd like you to read the article and answer some questions about it. I hope that I can get you to agree with me



Melissa Delgado

Hello, I'm Melissa Delgado. I wrote a letter to Students and Health magazine about Emil Fedor's article. I disagree with him, because I think schools should have junk food. Let's see if you'll agree with me after you read my letter.

To listen to Emil and Melissa speak, click **Play audio only**.


To listen and read at the same time, click **Play audio with text**.




Quick check for understanding: Purpose for Reading



Listening. Learning. Leading.®

Purpose for Reading








Now read the sentence below. Click on an answer to complete it.



The purpose of reading Emil Fodor's article and Melissa Delgado's letter is to _____.

- learn about where our food comes from
- decide which author's opinion we agree with
- find out what students like to eat around the world

Figure 3. Screenshots of the sample tasks in the warm-up and introductory section.

5. Research Questions

Upon the development of the assessment tool, we undertook a small-scale usability study to examine the extent to which the tool was utilized for intended formative purposes. This usability study was intended not only to collect validity evidence but also to inform the areas of further modification of the tool. Based on the formative assessment framework seen in Figure 1, we formulated research questions regarding the quality of assessment tasks and the roles of teachers and students in using the tool. Specifically, we posited the following research questions:

1. Are there differences in EL students' performance by subconstruct?
2. How do teachers implement the computer-based reading assessment tool for formative purposes? How do teachers interpret assessment results and use them for instructional adjustments for EL students' reading skills?
3. How do EL students perceive learning goals and feedback in a formative assessment process?
4. How do teachers perceive the usefulness of the computer-based formative assessment program? How do teachers perceive the quality of the formative assessment tasks in the tool for the targeted reading comprehension construct?

The first research question was concerned with the quality of assessment tasks. We were particularly interested in examining our underlying assumption that there would be a linear pattern of students' performance by subconstruct (i.e., foundational skill tasks would be easier than those assessing higher-order comprehension skills). The other research questions were also anticipated to provide insights into factors associated with the usefulness of the assessment tool for formative purposes.

6. Method

We employed a multiple case study approach (Baxter and Jack 2008) aimed at exploring differences and similarities in implementing formative assessment across different contexts. As a small-scale, exploratory study, this design enabled us to closely examine how individual teachers and students utilized our tool for formative assessment purposes. Below we describe the participants, study instruments, procedure, and data analysis.

6.1. Participants

In order to try out the present prototype materials in various settings, we recruited three ESL teachers from three secondary level schools. The teachers were teaching their EL students in different settings: (1) self-contained class (i.e., an ESL teacher teaching English language arts and English language development to a class of EL students), (2) push-in/co-teaching (i.e., an ESL teacher with a content teacher in an English class comprised of both EL and non-EL students), and (3) pull-out (i.e., an ESL teacher with a small group of EL students pulled out from mainstream English language arts classes).

In regard to students, a total of 62 EL students participated in this study with 55% students being female. Of the EL students, 89% spoke Spanish as their home language, while six different home languages were spoken among the remainder. Their English language proficiency ranged from low-intermediate to intermediate levels based on results on state English language proficiency assessments. Table 2 presents the class characteristics and the background of teachers included in this study.

Table 2. The characteristics of participating teachers, students, and class settings.

Teacher ID (Pseudonym)	Number of Years Teaching	Instructional Setting	Grades	Number of Participating Students	Student English Language Proficiency Levels	Other Languages Spoken by Teacher
Sue	10	Self-contained ESL/English language arts	7	19 ELs	Low-intermediate to intermediate	Spanish
Paula	13	Push-in/co-teaching English language arts	8	30 ELs, 16 non-ELs	Intermediate to high-intermediate	None
Kris	21	Pull-out ESL/English language arts	7–11	13 ELs	Low-intermediate to intermediate	Polish

6.2. Study Instruments

We developed multiple study instruments to gather diverse data sources. The instruments included: (a) a classroom observation protocol; (b) a teacher interview protocol; (c) a teacher survey; (d) a teacher reflection form for lesson planning; (e) a student background questionnaire; and (f) a student survey. A brief description of each instrument is included below.

Classroom Observation Protocol. This protocol was utilized by lesson observers (researchers). The protocol included instructions to take notes about how the teacher set up the lesson, how the teacher implemented the tool, how the students engaged with the tool, how the teacher and students used the assessment results, how the teacher interacted with the students, how the students interacted with their peers, and problems and challenges teachers and students faced while using the tool. The protocol also included guidance on what to include as concrete examples (e.g., classroom discourse).

Teacher Interview Protocols. Pre- and post-lesson observation interview protocols were employed to interview the teacher each day. The interview questions centered on teachers' use of the tool as well as their feedback on the tool (e.g., how the teachers were planning to use the formative assessment tool, how the teachers used assessment results to plan instruction, level of engagement of the students, the feedback that was provided to the students, how the students used the feedback).

Teacher Survey. The survey was also used to gather additional information about the teachers' teaching experience, their class, and their students. The survey questions also covered teachers' perceptions of the assessment tool, learning goals, assessment tasks, feedback, performance results, and resources for students (e.g., plan for improvement, additional activities, self-assessment, reflection form).

Teacher Reflection Form. This form was specifically designed to examine how teachers reflected on the evidence they gathered about their students' reading skills and how they were planning to use this evidence to guide their instruction. Teachers also reflected on how they interacted with their students (e.g., questions the teachers asked, questions the students had) and what they did to support their students. Teachers were asked to complete this form at the end of each lesson.

Student Background Questionnaire. This questionnaire was completed by the teachers, asking about students' background information (e.g., grade, gender, age, home language, English language proficiency scores and levels).

Student Survey. The study survey was designed to gather students' perceptions of the learning goals, activities, immediate feedback, performance results, plan for improvement, and additional resources presented in the tool.

6.3. Procedure and Analysis

The teachers attended a training session where the project goals and the prototype assessment tool were introduced. They also had an opportunity to navigate the tool on their own computers during the training session. The teachers were also asked to plan how to incorporate the tool in their classroom. The teachers used the tool during three class periods (90 min each) over a week. A pair of trained researchers observed each lesson. The researchers took detailed notes based on the observation protocol. On each day of the

observation, each teacher participated in two interviews, one prior to the lesson and one immediately afterward. All interviews were audio-recorded and transcribed. Regarding students' performance, the students' responses on the selected-response questions were machine-scored instantly and constructed-response questions were scored by teachers. Based on the scores, a report was generated and accessible to teachers and students within the tool's 'performance result' section. During the lesson, the teachers helped students review the results and complete all of the activities in the 'next steps and resources' sections (plan for instruction, final self-assessment, and student survey). The students' performance data were also collected for analysis.

All qualitative data, including classroom observation notes, interview transcripts, teachers' reflection forms, and open-ended survey responses, were coded by a pair of researchers, using a coding scheme we developed. The coding categories attempted to capture evidence about teachers' communication of learning goals, interpretation of learning evidence, feedback activities, lesson adjustment based on assessment evidence, perceptions about the assessment tool, and technology use. After independent coding, the two researchers met to compare their coding and reached agreement through discussion. For student survey and student performance data, descriptive statistics were computed.

7. Results

7.1. The Quality of Assessment Tasks

Although the validity and effectiveness of formative assessment lies in the assessment process, the quality of assessment tasks should not be neglected. In addition to the reliability coefficient (Cronbach's $\alpha = 0.856$), we examined the extent to which three subconstructs' items discriminated students' reading subskills. Figure 4 exhibits students' average percent correct on the items of each subconstruct reading skill. An expected performance pattern was observed in which the difficulty of items increased by subconstruct, in the order of foundational, literal comprehension, and higher-order comprehension skills. Interestingly, a wide range of performance variation was also observed in the higher-order reading skills compared to foundational and literal comprehension skills.

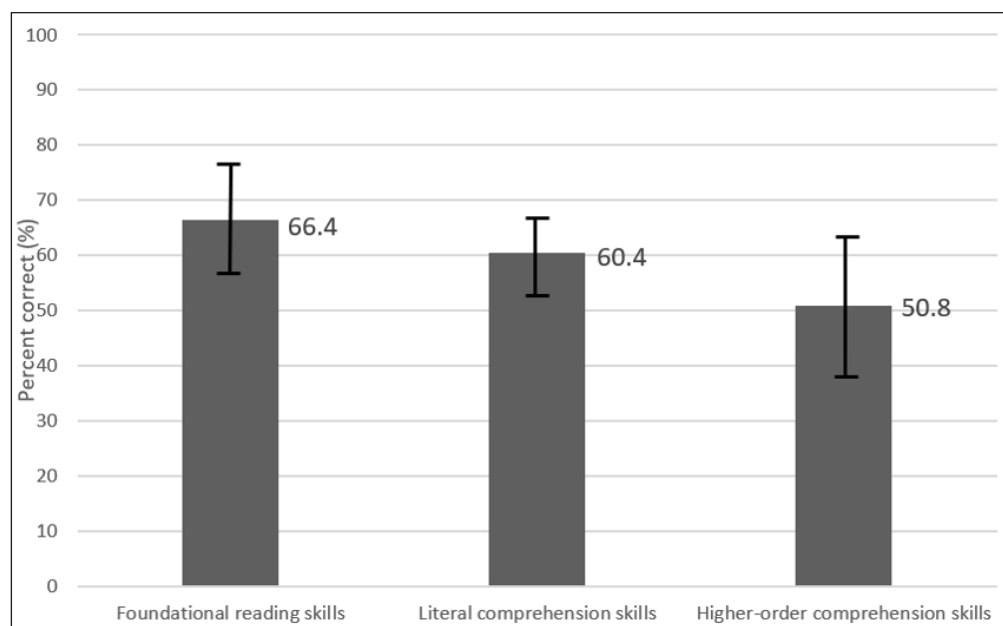


Figure 4. The average percent correct of students' performance on each subconstruct reading skill.

While the student score data provided some evidence about the quality of reading tasks to distinguish student levels in terms of three subskill levels, teachers' comments also testified to the quality of assessment tasks. All three teachers remarked that the

reading tasks were well-designed and covered important reading skills necessary for comprehending argumentative texts. For example, Sue commented, “I think every one of those skills is important. Everything that you had in there [in the tool] is things that we talk about, are things that the kids need”. More importantly, teachers provided specific examples regarding their interpretations of students’ reading skills based on individual tasks, indicating the quality of the given reading tasks for identifying students’ current status related to the target construct. This point is further described in the next section as part of teacher’s implementation skills.

7.2. Teachers’ Implementation of Formative Assessment

In this section, we summarize how three teachers (Sue, Paula, and Kris) utilized the tool for formative assessment purposes, based on our coding of the qualitative data. We organized the summaries with respect to three main elements of our formative assessment framework.

7.2.1. Setting Up and Communicating Learning Goals

All three teachers began with the first component (the learning goal) of the tool with students as a whole class. Teachers read aloud the learning goal including the reading standard language verbatim to students. However, there seemed to be little efforts to communicate the learning goal to students besides reading the text aloud. Only Sue asked students about the meaning of an argument stated in the learning goal and checked students’ understanding about what they were expected to learn over the next few class periods.

7.2.2. Collecting and Interpreting Learning Evidence

When students began to engage in assessment tasks (the activities component in the tool), teachers actively interacted with students. All teachers used at least one of the warm-up tasks to have students talk about their opinions on a given topic. This activity set up a context for students to state their own opinions as well as evaluate authors’ arguments in the reading passages. All three teachers closely monitored students’ engagement with the tasks during each lesson. During each post-lesson interview, teachers mentioned their assessment of students’ reading skills based on their interaction and observation. Teachers’ remarks indicated their formative assessment practice of interpreting students’ status based on evidence. For instance, Kris said, “. . . I interacted with all of my students throughout part 1. Sometimes I felt they were not paying enough attention to grammar or specific vocabulary words [in the reading passage]. I realized that a lot of the kids have a hard time with distinguishing fact or opinion. . . . They don’t understand like why it is a fact”. Paula also pointed out specific reading tasks to identify the areas students needed to improve, saying, “So the verb, nouns like um . . . and that’s something that ELs always struggle with . . . like the suffix, *tion* when to add the silent e. I noticed even the non ELs struggling with that grammatical structure and those foundations”. Paula further commented on students’ higher-order comprehension skills based on their associated tasks, “Today’s activities took longer than anticipated. The students were still not applying the skills taught the way that I would like them to use. Many of them were not citing textual evidence from the stories. That is a skill that we have been working on since the beginning of the school year”. Kris made similar comments in reflecting students’ performance, “They had some difficulty with the argumentative questions. Also, I found the questions that had suffixes added were confusing to my students. Inferencing questions were also hard to my students”. Notably, all three teachers referred to the task type names (e.g., working with words, facts and opinions, drawing inferences) in interpreting students’ performance and reading skills. It was evident that the task name attached to the targeted reading skill helped teachers make easy and quick interpretations of student performance.

7.2.3. Using Evidence: Providing Feedback and Planning Instruction

Teachers differed in the degree to which they adjusted their lessons based on student performance, either in real time or on the following day. For example, both Sue and Paula walked around their classes and noticed that many students struggled with the questions about facts vs. opinions as described in the previous section. Paula paused students' pair work and conducted a whole-class discussion by providing a few examples of factual and opinion statements during the lesson. This was a way of providing feedback to the whole class by making an immediate lesson adaptation based on real-time evidence. Sue made some quick dialogues with students about the definition of a fact and an opinion. On the other hand, Kris tended to focus on helping students to complete the tasks.

For the next-day lesson adjustment, Paula was the only teacher who acted on the results from the first lesson. She began the second lesson by modeling a response of some tasks that she found many students struggled with (e.g., writing reasoning). Sue also found some specific tasks that students had difficulty with. During the post-observation interview, she commented, "So yesterday, when I was scoring those seven students' responses, they really had a hard time with the one . . . finding two examples of evidence that supports this statement. I feel like most of them did not get that. And going back into the text like finding evidence like they would struggle". However, Sue did not make any adjustments to her second day lesson, focusing instead on students' completion of the remaining tasks. Similarly, Kris identified students' reading skills based on their performance as presented in the previous section. However, there was no lesson adaptation based on her interpretations.

Specificity about lesson plans based on the overall performance results was also varied among teachers. For example, Kris remarked, "Yes I would make specific instructional plans from the tasks they had trouble with using the activities in the Next Steps portion". However, there was no elaboration on how she would make such plans. Paula had more specific interpretations and plans, saying, "What I noticed after looking at many students' work, was that students seemed to struggle with inference questions, and questions pertaining to rephrasing main ideas. I could use the tools given to work with small groups to give further instruction. The students could then use the reference book [resources provided by the tool] to continue practicing on skills where they are weak. I would review making inferences with the whole class. I would then work with a small group of students that struggled the most to continue their instruction and help them better understand the material". Sue made a similar plan by stating, "When I teach in future lessons, I will limit the time I spend in review. I will use more of the 10-min mini lesson approach and move into the student activity sooner. I will definitely add more grammar and vocabulary to my lessons, as these were areas where my students need more reinforcement".

All teachers articulated future lesson plans based on both their real-time observation of students' performance and the overall performance results from the tool. Yet, their lesson plans appeared to focus on the types of tasks to work on rather than the next steps toward the overall learning goal.

7.3. Students' Perceptions

Central to formative assessment is articulating learning goals that are clear to the students, providing ongoing feedback and supporting students to engage in self-assessment and goal setting. In this section we provide findings from the survey to indicate how students perceived the learning goals, immediate feedback, and information in the performance results section as feedback.

Figure 5 provides information about how students perceived the clarity and usefulness of the learning goals and the initial self-assessment (can-do statements) for understanding the learning goals stated in the formative assessment tool. Although some stated that the learning goals (50%) and the initial self-assessment (26.3%) were clear, these were not transparent to all students. However, many students indicated that the learning goals were useful in helping them understand the purpose of the activities and that the initial

self-assessment was useful in helping them think about what they needed to do to complete the activities.

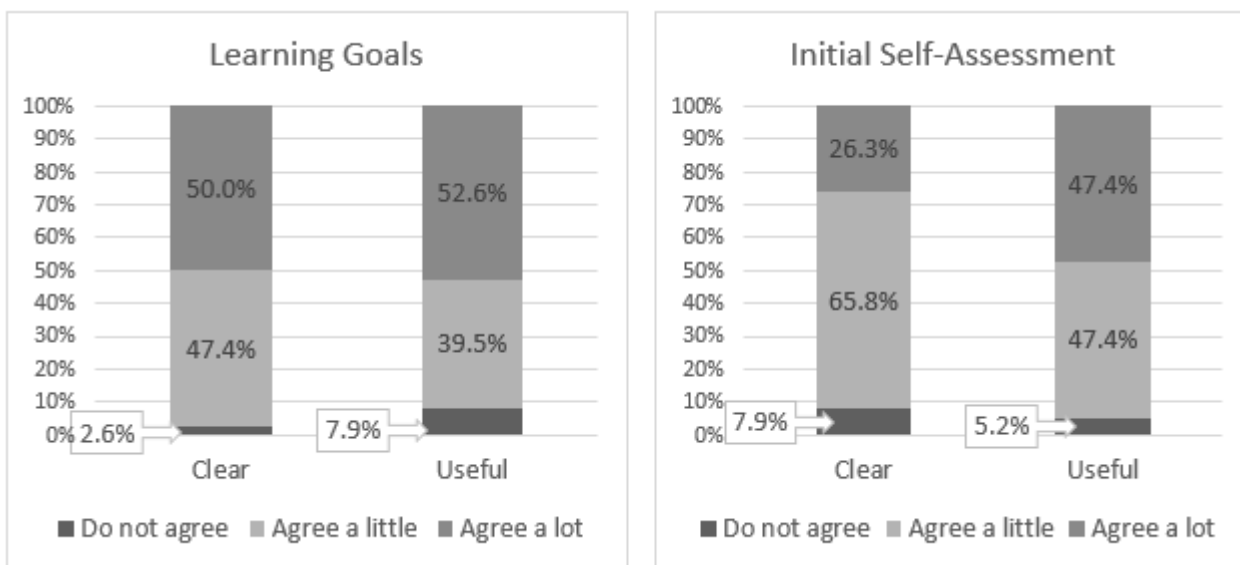


Figure 5. Clarity and usefulness of learning goals and initial self-assessment for understanding learning expectation for students.

In terms of the immediate feedback that was provided to the students while they completed the activities in part 1, half of them indicated in the survey that the feedback was clear and helpful in completing the activities (see Figure 6). However, there were a few students that indicated that the feedback was not clear (5.3%) or helpful (10.5%).

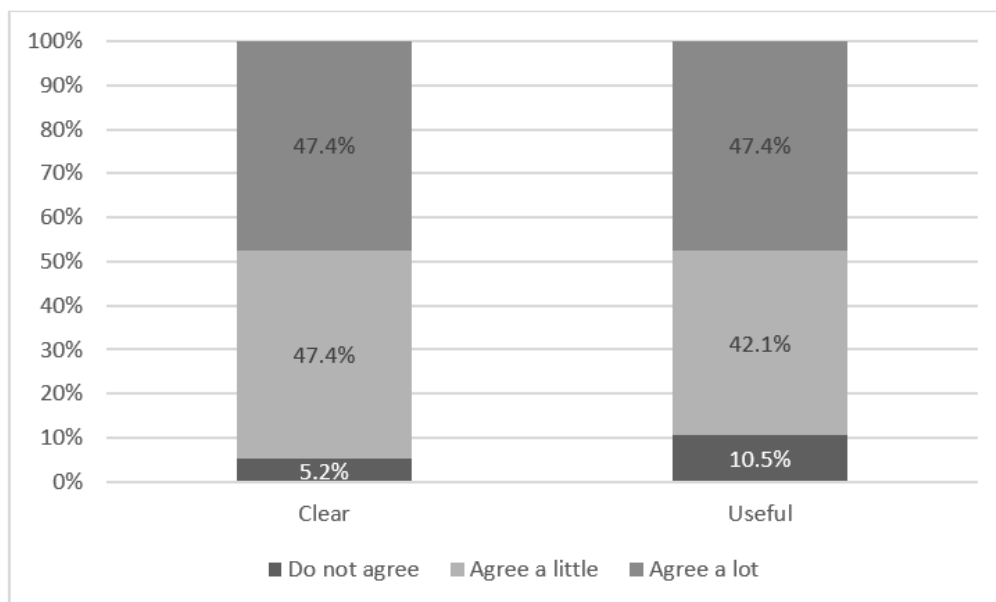


Figure 6. Clarity and helpfulness of immediate feedback for completing activities for students.

Students also indicated that the information in the performance results section (student responses, correct responses, items scores, skills scores, total scores, feedback) was useful in helping them understand what reading skills they had developed well (36.8%), what reading skills they still need to work on (31.6%), and what plans to set for improving specific reading skills (36.8%). The survey findings also suggest that many students still needed support in interpreting and using their assessment results (see Figure 7).

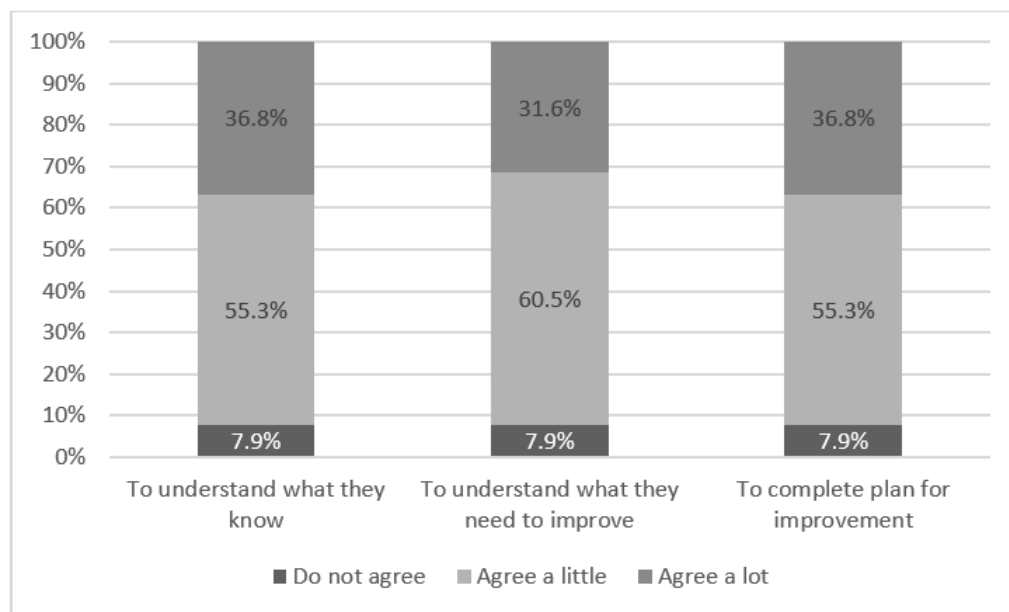


Figure 7. Usefulness of performance results in guiding learning for students.

7.4. Teachers’ Perceptions about the Usefulness of the Tool and the Quality of Assessment Tasks

In this section, we summarize the three teachers’ comments on the usefulness of the assessment tool, particularly regarding assessment tasks, specific features they found useful, and potential uses of the tool.

In general, teachers provided positive feedback on this prototype tool largely for three reasons. The first reason was that the assessment tasks encompassed key skills that students would need for reading argumentative texts. This feature appeared to help teachers plan their lessons addressing individual EL students’ needs. The tasks were deliberately designed for EL students to engage in from the foundational to higher-order reading skills with some scaffolding embedded. Sue summarized the overall feature of the tasks as in the following, “I felt, as a group, they [students] had a good understanding of what was being argued in each text. I believe this tool almost forces the students to think about what they read by asking various questions that require them to go back to the text to look for details and examples. By the time they finished these activities [reading tasks], each student had read each paragraph more than once, so I’m sure that helped with their comprehension skills”. Paula also explained how this tool allowed her to assess her students reading skills, “. . . It tells you like what they understood and what they did not understand. What I probably need to go back and reteach. What I need to focus on, what they had mastered also, so what I don’t need to spend time on. So, it does help gauge instruction”. In addition, the teachers commented that the content of the activities in the tool was generally appropriate, even though there were students at different language proficiency levels in their classrooms.

Secondly, teachers valued EL-specific design features. In particular, they pointed out the immediate feedback feature (e.g., “try again”), the collaborative tasks, and the special supports and tasks for EL students (e.g., Buzzy the Bee reading aloud the directions to the students, modeling what students need to do to complete an activity, highlighting the words students did not know in the text, tasks attending to word formation and sentence structures). Regarding collaborative tasks, Sue explained that her students were accustomed to working in pairs and that the tasks in this assessment tool allowed for peer collaboration. She commented that in classes with mixed English language proficiency levels, ESL teachers were constantly strategizing grouping; for example, she typically paired a lower-English proficiency level student with a higher-English proficiency level student so they can support each other in their home language. She mentioned that low-level students would benefit the most since they could ask their partners questions whenever they had

difficulties or were unsure of what they needed to do. She commented that her students were more willing to try to complete the tasks in small group environments than in a whole class setting. Kris also valued the pair work by stating, “I liked the pair work because it kept them [the students] on task. You know they had to have discussions in order to answer it. And it kind of allowed them to monitor each other instead of just clicking through it. It held them accountable”.

The third reason was attributed to technology features. Teachers asserted that their students were highly engaged in the tool due to its presentation (e.g., visuals, read-aloud/text-to-speech, different task formats) and students’ interactions with the tool (e.g., clicking, highlighting, navigating across different sections, receiving immediate feedback, and viewing score reports at the end of the lesson). Sue said, “. . . The activities are fun and engaging. My students were into it [the assessment tool]. They did not feel they were completing a test”.

Interestingly, observation notes indicated that teachers’ technology skills were varied and had an impact on the ways that teachers adapted the lessons in real-time. For example, Paula who was skillful at navigating the tool and the computer projector was able to switch whole-group discussions and students’ pair work more seamlessly than other teachers during each lesson. On the other hand, Sue (who acknowledged her limited computer skills) had difficulties in modeling examples based on the given tasks to the class using the program and computer projector, even though she noticed that many students struggled with a similar task. Teachers also pointed out potential technical issues such as the limited laptop availability and unstable internet connectivity in school.

In response to the questions about the potential usage of the tool, teachers indicated that this tool could be used in multiple ways in addition to formative assessment purposes. Their responses included the use of the tool as a main instructional material (i.e., curriculum unit), a mini-lesson material, a supplemental material for small-group work, and an end-of-unit test.

8. Discussion

In this study, we developed a technology-based assessment tool to help ESL teachers carry out formative assessment practices in efficient and systematic manners. Through a usability study of the prototype tool, we examined the ways in which the tool was used by teachers and students for the intended formative assessment use. The results of the usability study provided certain evidence to validate the tool for intended purposes.

Although the design of the tool contained four elements to reinforce the formative assessment process, teachers’ implementation skills were inconsistent. In general, the teachers in this study exhibited proficiency at analyzing and interpreting students’ performance and assessment results, as exemplified in the results section. Considering the prominent use of standardized assessments and accountability data reporting for the past two decades in U.S. K-12 education, the teachers were familiar with analyzing the student data. However, some limited implementation skills were noted in communicating learning goals to students. The lack of communication about the learning goals seemingly narrowed teachers’ immediate and near-term lesson planning only to the types of activities and tasks the students would complete. Teachers’ reflection on students’ performance centered on specific tasks and their associated reading skills (e.g., vocabulary, inferencing). That is, teachers’ interpretation of learning evidence focused primarily on the current status, rather than the gap between the current status and the overall learning goal. This tendency led teachers to mainly articulate the types of tasks to work on as next steps, neglecting plans based on individual EL students’ performance relative to the learning goal. This finding suggests that teachers would benefit from further professional development on the concept of formative assessment. In distinguishing formative assessment from summative assessment, [Heritage \(2013\)](#) points out that teachers need to develop “a present-to-future perspective, in which the concern is not solely with the actual level of performance but with

anticipating future possibilities” for realizing the intent of formative assessment (Heritage 2013, p. 180).

In addition, we observed limited teacher feedback. As presented in the results section, one teacher was more skillful at providing real-time feedback and making immediate lesson adaptation than the other two teachers. Overall, there were few classroom conversations where the teacher asked further probing questions contingent upon individual students’ responses on the given task in real time. The dialogues were primarily limited to completing the tasks and classroom management. The literature on formative assessment stresses a variety of methods of formative assessment including teacher questioning. Ruiz-Primo and Furtak (2006) called them assessment conversations and provided empirical evidence of the positive relationship between the quality of assessment conversations as a formative assessment method and student learning outcomes. While we intended our tool for teachers to practice formative assessment processes, teachers appeared to use it more as a summative assessment tool, focusing on students’ completion to review the complete assessment results at once. Teachers’ comments about using the tool as a unit-test or supplementary material to assign to small group work align with the finding about teachers’ limited assessment conversations in this study. This finding about teachers’ misconceptualization about their use of summative assessment as formative assessment is also found in previous research (e.g., De Lisle 2015; Heritage et al. 2009). In De Lisle’s (2015) study, for example, teachers were skillful at recording and storing data, but neglecting the use of data, which was the essential aspect of formative assessment. As Saito and Inoi (2017) note based on their study, teachers’ clear understanding about the purposes of formative assessment is crucial for the effective implementation of formative assessment. For further development of an assessment tool, it may be worth including some examples of probing questions and a vignette to describe the use of the tool for formative purposes. This addition may help raise teachers’ metacognitive awareness about the intended use of formative assessment.

We deliberately designed the tool to involve students in the formative assessment process (e.g., initial self-assessment, post-activity self-assessment upon reviewing the performance results, feedback during and after the activities). It was interesting to find that students found the learning goals useful to know, even though the clarity of the learning goals remains an area for improvement. Students’ English language proficiency levels and data interpretation skills seemed to hamper them from understanding the feedback in the performance result section and planning on next steps. While the tool has room for improvement, the finding was promising for students to take an agentive role in understanding learning goals, reflecting on their performance, and feedback.

Although the study is limited in the small number of teachers with a short period of lesson observation, the study findings offer useful implications for practice and further research. We acknowledge that teachers’ limited use of the tool for formative assessment purposes might have to do with their lack of understanding the intent of the tool. Thus, the findings should be interpreted with caution. Yet, our three teachers’ varied and limited implementation skills to use the tool for formative assessment corroborate previous research findings about the importance of providing professional support to increase teachers’ language assessment literacy skills, particularly for the practice of formative assessment (Fulcher 2012; Heritage 2007; Leung and Scott 2009; Tsagari 2016). In current U.S. K-12 education settings, a heavy reliance on large-scale standardized assessments for accountability appears to result in a focus of professional development on summative assessment. While formative assessment has gained traction to balance the assessment system, professional support to increase teachers’ capacity to realize the benefit of formative assessment warrants an increased emphasis.

It is also important to recognize teachers’ heavy workload and to have realistic expectations of teachers. Teachers constantly juggle tasks with limited time and resources while ensuring coverage of the curriculum for all students. For ESL teachers who have to deal with both English language proficiency and academic content standards to support EL students’ academic success in mainstream classes, we argue that it is critical to provide a

sound tool to alleviate teachers’ burden to devise methods to implement good formative assessment practice in a sustained way. The tool developed in this study based on a theoretical framework provides one example. Although there was a limited use of the tool for formative assessment in this study, the study findings revealed sources of misuse and areas of further investigation. We call for further research on the development of a tool for formative assessment and usability studies to inform best practices of utilizing such a tool for effective formative assessment. Future research should also investigate student learning outcomes as well as agents’ (teachers and students) behavioral changes that contribute to effective formative assessment practice.

Author Contributions: Conceptualization, M.K.W.; Formal analysis, M.K.W. and A.A.L.; Investigation, M.K.W. and A.A.L.; Supervision, M.K.W.; Writing—original draft, M.K.W. and A.A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Committee for Prior Review of Research (CPRR) and approved by CPPR. The CPPR is Educational Testing Service’s Institutional Review Board.

Informed Consent Statement: Informed consent was obtained from all teachers involved in the study. For students involved in the study, their parent/guardian consent was obtained accompanied by student assent.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank Michael Ecker, Chris Hamill, Keith Kiser, Maria Konkel, Nathan Lederer, Jeremy Lee, Janet Stumper, and Jennifer Wain for their helpful research assistance in developing the tool and collecting the data for this study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Part 1 Sample Tasks

The screenshot displays a digital reading task interface. At the top, there is a navigation bar with the ETS logo and the title 'Read for Key Words: Question 1'. Below this, there are two tabs for 'Article 1' and 'Article 2'. The main content area shows the text of 'Opinion: No More Junk Food in Schools' by Emil Fedor. The text is divided into three paragraphs. Various words are highlighted in red, and a list of related words is shown in a pop-up box. For example, the word 'diseases' is highlighted, and a list of related words includes 'dangers', 'sicknesses', and 'accidents'. Another example shows the word 'banned' highlighted, with a list of related words including 'not allowed', 'celebrated', and 'read'. A third example shows the word 'angry' highlighted, with a list of related words including 'colorful' and 'a lot'. To the right of the text, there are three questions (Q1a, Q1b, Q1c) and a bee character. Q1a asks the user to choose a word that is most similar in meaning to the one in red. Q1b shows an example: 'She has a very large backpack.' with a list of related words: 'big', 'dry', and 'yellow'. Q1c asks the user to choose a word that is most similar in meaning to the one in red. The interface also includes a search bar, a magnifying glass icon, and a 'Click Next when you're finished.' button.

Figure A1. Sample task 1 for the subconstruct of foundational skills (Task type: Working with words).

Figure A2. Sample task 2 for the subconstruct of literal comprehension skills (Task type: Getting the details).

Figure A3. Sample task 3 for the subconstruct of literal comprehension skills (Task type: Distinguishing facts from opinions).

Working with Words: Question 12

Now, fill in the empty boxes in the Verb and Noun columns with the correct form of each word. Then fill in the blank to the right with the correct form. The first one has been done for you. When you're finished, click **Show me the answers!** to check your work.

Verb	Noun	Sentences
educate	education	Part of a student's <u>education</u> should be learning how to eat right.
participate	Q12a	Q12b Schools around the country _____ in Health Awareness Month.
Q12c	<u>promotion</u>	Q12d In order to _____ healthy eating, some schools have removed their soda machines.
imitate	Q12e	Q12f Sometimes you may think that you are drinking real orange juice, but it is only an _____.
Q12g	<u>digestion</u>	Q12h Foods with a high fat content can be difficult to _____.

Show me the answers! Click **Next** when you're finished.

Figure A4. Sample task 4 for the subconstruct of foundational skills (Task type: Working with words).

Working with Argument Structure: Question 19

Now reread the article. After reading, fill in the graphic organizer from the answer choices (A, B, C, D, or E).

Answer Choices

- A. There are many health problems associated with eating junk food.
- B. Some people think students should eat whatever they want.
- C. A school cafeteria in New York City sells only healthy and delicious vegetarian lunches.
- D. Students will think eating junk food is OK if school cafeterias sell it.
- E. Vegetables, fruits, and grains are better for you than junk foods.

Graphic Organizer:

- MAIN IDEA Paragraph 1:** Junk food is not healthy and it should not be sold in school cafeterias.
- REASON Paragraph 2:** The first reason is because _____ (Q19a)
- REASON Paragraph 3:** The second reason is because _____ (Q19b)
- EXAMPLE TO SUPPORT MAIN IDEA Paragraph 4:** A good example is _____ (Q19c)
- CONCLUSION Paragraph 5:** Schools need to do the right thing—stop selling junk food and only sell healthy food to students.

Click **Next** when you're finished.

Figure A5. Sample task 5 for the subconstruct of higher-order comprehension skills (Task type: Working with argument structure).

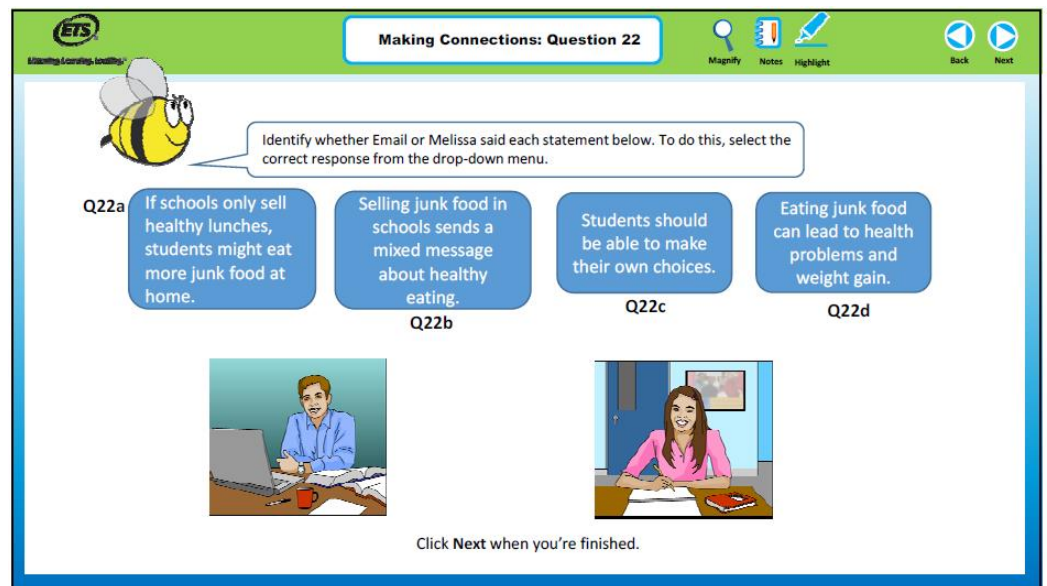


Figure A6. Sample task 6 for the subconstruct of higher-order comprehension skills (Task type: Making connections).

Appendix A.2. Part 2 Sample Task

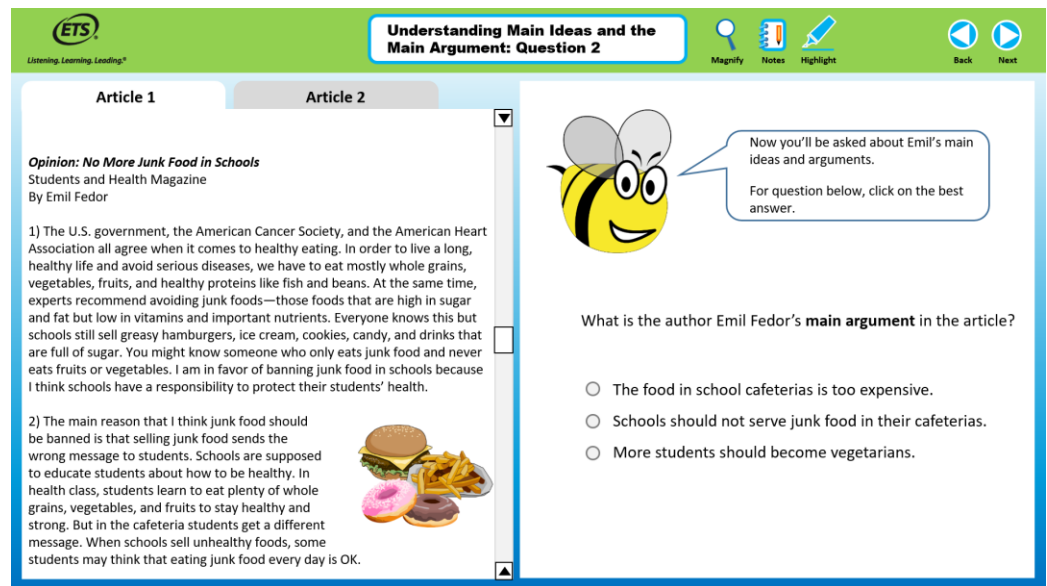


Figure A7. Sample item for the subconstruct of literal comprehension skills (Task type: Understanding main ideas).

References

Alderson, J. Charles. 2000. *Assessing Reading*. Cambridge: Cambridge University Press.

Bailey, Alison L., and Mikyung Kim Wolf. 2020. The construct of English language proficiency in consideration of college and career readiness standards. In *Assessing English Language Proficiency in U.S. K-12 Schools*. Edited by Mikyung Kim Wolf. New York: Routledge, pp. 36–54. [CrossRef]

Baxter, Pamela, and Susan Jack. 2008. Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report* 13: 544–56. [CrossRef]

Bennett, Randy E. 2011. Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice* 18: 5–25. [CrossRef]

Bernhardt, Elizabeth B. 2011. *Understanding Advanced Second-Language Reading*. New York: Routledge.

Black, Paul, and Dylan Wiliam. 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice* 5: 7–74. [CrossRef]

- Black, Paul, and Dylan Wiliam. 2010. Inside the black box: Raising standards through classroom assessment. *Kappan* 92: 81–90. [CrossRef]
- Bunch, George C. 2013. Pedagogical language knowledge: Preparing mainstream teachers for English learners in the New Standards Era. *Review of Research in Education* 37: 298–341. [CrossRef]
- Callahan, Rebecca M. 2013. *The English Learner Dropout Dilemma: Multiple Risks and Multiple Resources*. California Dropout Research Project Report #19. Santa Barbara: University of California.
- Common Core State Standards Initiative. 2010. *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Available online: http://www.corestandards.org/wp-content/uploads/ELA_Standards1.pdf (accessed on 2 May 2021).
- Council of Chief State School Officers. 2014. *English Language Proficiency (ELP) Standards with Correspondences to K-12 English Language Arts (ELA), Mathematics, and Science Practices, K-12 ELA Standards, and 6–12 Literacy Standards*. Washington, DC: Council of Chief State School Officers.
- De Lisle, Jerome. 2015. The promise and reality of formative assessment practice in a continuous assessment scheme: The case of Trinidad and Tobago. *Assessment in Education: Principles, Policy & Practice* 22: 79–103. [CrossRef]
- Every Student Succeeds Act. 2015. 20 U.S.C. § 6301. Available online: <https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf> (accessed on 3 September 2021).
- Fulcher, Glenn. 2012. Assessment literacy for the language classroom. *Language Assessment Quarterly* 9: 113–32. [CrossRef]
- Gan, Zhengdong, and Constant Leung. 2019. Illustrating formative assessment in task-based language teaching. *ELT Journal* 74: 10–19. [CrossRef]
- Grabe, William. 2009. *Reading a Second Language: Moving from Theory to Practice*. Cambridge: Cambridge University Press.
- Hamill, Christopher, Mikyung Kim Wolf, Yuan Wang, and Heidi Liu Banerjee. 2019. *A Review of Digital Products for Formative Assessment Uses: Considering the English learner Perspective*. Research Memorandum No. RM-19-04. Princeton: Educational Testing Service.
- Heritage, Margaret, Jinok Kim, Terry Vendlinski, and Joan Herman. 2009. From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practices* 28: 24–31. [CrossRef]
- Heritage, Margaret. 2007. Formative assessment: What do teachers need to know and do? *Phi Delta Kappan* 89: 140–45. [CrossRef]
- Heritage, Margaret. 2010. *Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity?* Washington, DC: Council of Chief State School Officers.
- Heritage, Margaret. 2013. Gathering evidence of student understanding. In *SAGE Handbook of Research on Classroom Assessment*. Edited by James H. McMillan. Thousand Oaks: SAGE Publications, Inc., pp. 179–95.
- Koda, Keiko. 2004. *Insights into Second Language Reading: A Cross-Linguistic Approach*. Cambridge: Cambridge University Press.
- Leung, Constant, and Bernard Mohan. 2004. Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing* 21: 335–59. [CrossRef]
- Leung, Constant, and Catriona Scott. 2009. Formative assessment in language education policies: Emerging lessons from Wales and Scotland. *Annual Review of Applied Linguistics* 29: 64–79. [CrossRef]
- Leung, Constant. 2004. Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly* 1: 19–41. [CrossRef]
- O'Reilly, Tenaha, and Kathleen M. Sheehan. 2009. *Cognitively Based Assessment of, for, and as Learning: A 21st Century Approach for Assessing Reading Competency*. Research Memorandum No. RM-09-04. Princeton: Educational Testing Service.
- Popham, W. James. 2008. *Transformative Assessment*. Alexandria: Association for Supervision and Curriculum Development (ASCD).
- Rea-Dickins, Pauline. 2001. Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing* 18: 429–62. [CrossRef]
- Ruiz-Primo, Maria Araceli, and Erin Marie Furtak. 2006. Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment* 11: 237–63. [CrossRef]
- Ruiz-Primo, Maria Araceli, and Min Li. 2013. Analyzing teachers' feedback practices in response to students' work in science classrooms. *Applied Measurement in Education* 26: 163–75. [CrossRef]
- Ruiz-Primo, Maria Araceli, Guillermo Solano-Flores, and Min Li. 2014. Formative assessment as a process of interaction through language: A framework for the inclusion of English language learners. In *Designing Assessment for Quality Learning, The Enabling Power of Assessment*. Edited by Claire Wyatt-Smith, Valentina Klenowski and Peta Colbert. Heidelberg: Springer, vol. 1, pp. 265–82.
- Sabatini, John, Elizabeth Albro, and Tenaha O'Reilly. 2012. *Measuring Up: Advances in How We Assess Reading Ability*. Lanham: R&L Education.
- Saito, Hidetoshi, and Shin'ichi Inoi. 2017. Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly* 14: 213–33. [CrossRef]
- Schildkamp, Kim, Fabienne M. van der Kleij, Maaïke C. Heitink, Wilma B. Kippers, and Bernard P. Veldkamp. 2020. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research* 103: 101602. [CrossRef]
- Shepard, Lorrie A. 2009. Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice* 28: 32–37. [CrossRef]
- Sugarman, Julie, and Kevin Lee. 2017. *Facts about English Learners and the NCLA/ESSA Transition in California*. Washington, DC: Migration Policy Institute.

-
- Tsagari, Dina, and Karin Vogt. 2017. Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment* 6: 41–63.
- Tsagari, Dina. 2016. Assessment orientations of primary state school EFL teachers in two Mediterranean countries. *Center for Educational Policy Studies Journal* 6: 9–30. [[CrossRef](#)]
- U.S. Department of Education, Office of English Language Acquisition [OELA]. 2021. *English Learner Population by Local Education Agency Fact Sheet*; Washington, DC: OELA. Available online: https://ncela.ed.gov/files/fast_facts/20210315-FactSheet-ELPopulationbyLEA-508.pdf (accessed on 3 August 2021).