The CATESOL Journal

MELISSA LATHAM KEH
Bridgewater State University

Keeping Language in Mind: An Exploratory Study of English Learners' Performance on Three Language and Literacy Assessments

Assessing English learners (ELs) in US schools is challenging because many widely used assessments have not been designed with ELs in mind. Yet if teachers are sensitive to how ELs may perform differently from native speakers on such assessments, these assessments reveal useful information about ELs' language and literacy skills. This mixed-method study compared adolescent ELs' performance on the Qualitative Reading Inventory-5 (QRI-5), Peabody Picture Vocabulary Test-4 (PPVT-4), and Words Their Way Elementary Spelling Inventory (ESI) to existing data from English-proficient examinees' performance to explore how ELs' performance may differ. The observed differences suggest that linguistic aspects of the QRI-5, PPVT-4, and ESI, including syntax, phonology, orthography, and especially vocabulary, played a role in ELs' performance and indicate that ELs may benefit from linguistic modification or first-language support during test administration. The process used in this analysis also demonstrates how teachers can examine test data alongside test scores as they interpret ELs' results.

To effectively account for the language and literacy development of English learners (ELs) and to provide ELs with targeted instruction, it is essential to have valid assessments of their language and literacy skills. Although assessments used with ELs should undergo field testing with ELs, some assessments that are used in schools with ELs have not undergone such evaluation. García, McKoon, and August (2006b) point out a wide range of concerns that educators and researchers have raised about using language and literacy assessments that are not specifically designed for language-minority students with such learners in the *Report of the National Literacy Panel on Language-Minority Children and Youth.* One concern is

that language and literacy assessments normed for native English speakers may not be valid or reliable when used with nonnative speakers of English and ELs. In other words, when normal performance on a given assessment is determined with a group that does not include nonnative speakers or ELs, it cannot be known for sure what normal performance might look like for these learners. Another concern is that the linguistic demand of assessments is a major factor in ELs' performance (Abedi, 2002; Abedi & Gándara, 2006). To address these concerns, this exploratory study analyzes EL performance on three assessments, which have not been extensively examined with ELs, with a specific focus on the role of linguistic demands on assessment outcomes.

The Qualitative Reading Inventory-5 (QRI-5; Leslie & Caldwell, 2011), Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn, 2007), and Words Their Way Elementary Spelling Inventory (ESI; Bear, Invernizzi, Templeton, & Johnston, 2012) were chosen for this study because they hold potential for use with ELs. Peregoy and Boyle (2016) recommend informal reading inventories, such as the QRI, as a tool for assessing ELs' reading. The QRI is a well-respected informal reading inventory (Nilsson, 2008) that, beyond its use in this study, I have seen used in Massachusetts to assess elementary-level ELs in schools and a high school EL at a tutoring center. The PPVT-4 is a widely used assessment of receptive vocabulary (Umberger, 1985), and speech-language pathologists commonly learn to administer the most recent edition of the PPVT in graduate programs (Susan Gray, personal communication, January 29, 2018). Additionally, the PPVT was one of the most common language assessments that school-based speech-language pathologists from 34 states, including California, indicated using with bilingual students in a survey conducted by Arias and Friberg (2017), and it was the most commonly used vocabulary measure in the literacy research with second-language learners that was included in the National Literacy Panel's meta-analysis (García, McKoon, & August, 2006a). Many of the teachers I have taught in the Boston metropolitan area use *Words Their Way* (Bear et al., 2012) and *Words Their Way With English Learners* (Bear, Helman, Templeton, Invernizzi, & Johnston, 2007), which include the *Words Their Way* spelling inventories.

Together these three assessments can provide an overview of a student's reading skills (e.g., decoding, fluency, and comprehension), vocabulary knowledge, and spelling skills, respectively, that would aid teachers in planning targeted literacy instruction. For ELs, this is critical because these learners may need reinforcement of specific skills that are not covered in the general education curriculum at their grade levels (Helman, 2005). Additional information about each assessment is provided in the next sections.

QRI-5

The QRI-5 is an informal assessment designed to determine instructional reading level. The passages for the QRI were pilot tested with children from three university clinics who represented various racial and ethnic groups and have been subsequently tested as new passages were developed and added in the QRI-4 (Leslie & Caldwell, 2006) and QRI-5 (Leslie & Caldwell, 2011). The pilot samples did not include ELs, but the QRI-5's qualitative nature is in keeping with assessment recommendations for ELs, in that it can allow a test administrator to give ELs opportunities to demonstrate their reading skills and comprehension in a variety of ways. For example, allowing ELs to look back at the text when answering comprehension questions reduces the amount that ELs, who may also be thinking about how to formulate answers in their second language, must remember and think about. Additionally, the option of scoring miscues using either Total Accuracy or Total Acceptability analysis makes it possible to classify and evaluate grammatical errors (e.g., non-meaning–changing miscues such as repeated omission of morphemes *–ed* and *–s*), which may stem from ELs' interlanguage (Selinker, 1972) rather than reading skills (Latham Keh, 2017), differently from miscues that change meaning. The QRI-5 manual provides some guidance for miscue analysis with Spanish speakers, cautioning that determining whether one hears a pronunciation difference or a reading error is "problematic when teachers who use Standard American English are listening to children … whose pronunciation reflects the sound system of another language" (Leslie & Caldwell, 2011, p. 17). However, the authors do not provide in-depth guidance for miscue analysis of nonnative English speakers beyond this, and there are no special instructions for administering or scoring ELs' performance on other components of the QRI-5.

The present study will focus specifically on one of these other components—the concept questions, which are intended to measure prior knowledge. An accurate measure of prior knowledge is important because prior knowledge is an influential factor in reading comprehension (Dochy, Segers, & Buehl, 1999) for both native English speakers and ELs, and research has long indicated the potential for ELs to draw from different background experiences than those of the mainstream population (Jiménez, García, & Pearson, 1996). As part of pilot testing, Leslie and Caldwell (2006, 2011) examined correlation between concept questions and comprehension questions for QRI passages read by at least 10 students. Although they do not provide correlation coefficients for every text in the QRI-5, they report that, overall, there is a relationship between conceptual knowledge and comprehension, especially for older students. Results indicate that students who score 55% or higher on concept questions typically score at least 70% on

passage-comprehension questions after reading. Many of the concept questions on the QRI-5 follow the definitional free-association style that Leslie and Cooper (1993) found most directly revealed students' knowledge. A description of this style of question and how it may influence ELs' answers will be discussed in more detail below.

PPVT-4

The PPVT-4 provides a measurement of receptive vocabulary independent of literacy skills, and because it is designed for a large age range, the most basic vocabulary sets can be administered to older ELs and adults without concern that the design of the assessment is inappropriate for learners their age. Because ELs were not part of the normative sample, Dunn and Dunn (2007) recommend that the PPVT-4 be used only as a criterion measure for individuals who are not English proficient. In other words, the PPVT-4 should not be used to determine whether an EL is typical or normal compared to other individuals at the EL's age or grade level, but it could be used to measure how much receptive vocabulary an EL has, and scores could be looked at through time to see how an EL is progressing. Garcíaet al. (2006a) also indicate that the PPVT may appropriately estimate how well language-minority children's recognition of mainstream English vocabulary matches that of native English-speaking students, although they caution that language-minority children and ELs may score lower than native English-speaking children on the PPVT (Fernandez, Pearson, Umbel, Oller, & Molinet-Molina, 1992; Sattler & Altes, 1984).

ESI

The ESI is designed to determine a child's developmental stage of spelling knowledge and can help teachers identify and target specific phonic features that students need to learn. The *Words Their Way* primary, elementary, and upper spelling inventories have been found to be valid indicators of academic performance in English-speaking, general-education classrooms based on the California Standards Tests and, furthermore, to reliably differentiate between relatively higher- and lower-performing elementary and middle school students, including when these samples include students identified as English language learner (ELL), SPED, or gifted (Sterbinsky, 2007).[1] Yet, to my knowledge, this study did not disaggregate ELs to examine their performance as a subgroup, and it was unclear how many of the students in the study were ELs. This would be helpful because Bear et al. (2007) caution that ELs may exhibit inconsistencies on this assessment, at times spelling words correctly that are beyond their developmental spelling level. With this in mind, this analysis will explore the extent to which word frequency in Standard American English explains whether or not ELs spelled a word correctly on the ESI.

In the long term, if these assessments will be used on a large-scale basis with ELs, their psychometric properties should be examined with ELs as separate populations. In the short term, a mixed-method, exploratory analysis can help teachers who are already using these assessments to better understand how ELs' performance may differ from that of native English speakers. Specifically, this study will address the following research questions:

1. Is there a similar correlation between concept questions and comprehension questions on the QRI-5 for EL readers and the pilot sample of proficient English speakers?
2. Do ELs need to take the same number of vocabulary sets as the English-proficient normative sample during PPVT-4 test administration?
3. To what extent does word frequency explain ELs' spellings on the ESI?

The results and discussion reveal unique patterns in ELs' performance on the three assessments and highlight how linguistic factors, especially vocabulary knowledge, may have played a role in outcomes. Recommendations for practical use of these assessments with ELs are provided.

Method

Research Design

The data for this analysis stem from a previous study (Latham Keh, 2014) that examined the influence of two approaches, retrospective miscue analysis (RMA; Goodman & Marek, 1996; see also Goodman, Martens, & Flurkey, 2014) and word study, on ELs' reading. For both trial approaches, the QRI-5 was used to measure baseline and posttrial reading skills, and the PPVT-4 was used to measure baseline and posttrial vocabulary knowledge. For the word-study approach, the ESI was used as an additional baseline and posttrial measure for literacy skills because the word-study approach was designed to potentially affect spelling as well as reading. The current analysis focuses exclusively on the EL participants' performance on the three baseline measures: QRI-5, PPVT-4, and ESI.

Participants

The 11 participants were adolescent ELs attending a suburban high school in Massachusetts. Their English-language proficiency ranged from level 1 to level 4, as measured by the Massachusetts English Language Assessment—Oral (Massachusetts Department of Elementary and Secondary Education, 2010), which described five levels of English proficiency. The participants spoke a total of seven different languages. All participants had

arrived in the US after age 10, and most had been in the US for one year or less. None of the participants had experienced gaps in formal education, and all were literate in their native languages. None of the participants were known to have any disabilities or exceptionalities. All 11 participants were administered QRI-5 reading passages and the PPVT-4. Four participants were also administered the ESI (see Table 1).

Table 1
Participants

Participant	*Native language*	*English proficiency on MELA-0*	*Age at time of study*	*QRI-5*	*PPVT-4*	*ESI*
Participant 1	Haitian Creole	1	15	✔	✔	✔
Participant 2	Mongolian	2	16	✔	✔	
Participant 3	Bulgarian	2	15	✔	✔	✔
Participant 4	Mandarin	3	14	✔	✔	
Participant 5	Mandarin	3	14	✔	✔	
Participant 6	Korean	3	16	✔	✔	✔
Participant 7	Mandarin	4	15	✔	✔	
Participant 8	Russian	4	18	✔	✔	
Participant 9	Mandarin	4	17	✔	✔	
Participant 10	Japanese	4	16	✔	✔	✔
Participant 11	Bulgarian	4	16	✔	✔	

Materials

The QRI-5 was used to determine instructional reading level, and concept questions were used to gather information about the participants' prior knowledge about the texts they read. QRI-5 passages typically have three to five concept questions that are scored on a four-point scale (see Appendix A). Leslie and Caldwell (2011) report 98% agreement for interscorer reliability for a sample of scorers who used the instructions and examples in the manual. As a measure of predictive validity, correlations of concept-question scores to passage comprehension were compared to correlations of general measures of reading achievement to passage comprehension; concept questions correlated significantly with passage comprehension more frequently than the general measures.

The PPVT-4 was used to measure receptive vocabulary knowledge. Age norms for the PPVT-4 are based on a representative set of 3,540 English-proficient individuals between the ages of two and a half and 90. This in-

strument includes two versions (A and B) designed to be highly similar in content and found to have negligible average-item difficulties. Internal-consistency reliability, alternate-form reliability, and test-retest reliability indicate a high level of score consistency on both Version A and Version B of the PPVT-4. Validity was established through the vocabulary word–selection process, comparison of average performance on the PPVT-4 to crystallized ability, and correlations between PPVT-4 scores and scores on expressive vocabulary, language ability, and reading-achievement instruments (Dunn & Dunn, 2007). The ESI was used to measure spelling knowledge. Analysis of reliability with a sample of 862 students from a suburban school district in California included item difficulty, item discrimination, and internal consistency, and resulted in an overall reliability coefficient of .915 (Cronbach's alpha). Test-retest reliability coefficients ranged from .700 to .974. Although some of the predictive validity coefficients for the third grade were not significant, all other predictive validity coefficients and all concurrent validity coefficients for second through fifth grade were significant at the $p < .001$ level (Sterbinsky, 2007).

Procedure

The author obtained informed consent from parents and assent from participants. After this, the author administered the QRI-5, the PPVT-4, and (if participants were participating in the word-study trial) the ESI. The assessments were administered during participants' scheduled EL tutoring time or during free time in a classroom in their school. In most cases, assessments were administered on different days because of time constraints and to ensure that participants would not be fatigued by taking multiple assessments. All assessments were administered and scored by the author using standard administration and scoring procedures, unless otherwise noted below.

For the QRI-5, the participants were first asked the 3-5 concept questions. Next, they read the text aloud while miscues were recorded on a separate transcript of the passage. The oral reading was timed for fluency. After reading, participants were asked the comprehension questions at the end of the passage using the look-back option. QRI-5 concept questions were scored according to the four-point scale provided by Leslie and Caldwell (2011), and comprehension questions were scored 1 point for correct or 0 for incorrect. Miscues and fluency were scored according to standard procedures in the manual and were not used for the current analysis. Since readers may perform differently on different QRI-5 passages at the same level (because of prior knowledge and/or text type), participants each read 2-4 passages from the QRI-5 before and after the trial approaches. As a group, the participants read 51 QRI-5 passages. The titles and levels of the passages that individual participants read are indicated in Appendix B.

The PPVT-4 includes 19 sets of 12 vocabulary items, arranged to increase in difficulty as the examinee progresses. For each item, the examinee sees four pictures and is prompted to indicate which picture depicts an orally presented word. Typically, the administrator begins with the set that matches the examinee's biological age and progresses through sets until the examinee identifies the wrong picture for eight or more words in a set. Because the participants in this study were ELs, the author usually did not begin with the PPVT-4 set corresponding to the participant's biological age. Instead, vocabulary sets were previewed, and testing began with a set the author believed would result in a basal score (no more than one word wrong) for the participant. If the participant got more than one word wrong in the first set, the next easiest set was administered until a set resulted in a basal score, and then increasingly harder sets were administered until a set resulted in a ceiling score (eight or more words wrong). Participants were administrated different versions (A and B) for pre- and posttesting to ensure they were equally unfamiliar with test items each time (see Table 2), and standard scoring procedures were followed.

The ESI was administered before and after the word-study trial approach only, as a measure of the four participants' spelling skills. Participants were instructed to number a blank piece of paper from one to 25. The words and sentences containing the words were read aloud to the participants, and the participants spelled the words on their paper. Participants were asked to try to spell all 25 words, and standard scoring procedures were followed.

Data Analysis

Quantitative analysis was used to compare the data from the ELs' performance on these three assessments to existing data from validity and reliability testing (Dunn & Dunn, 2007; Leslie & Caldwell, 2006, 2011; Sterbinsky, 2007). A Spearman rank correlation was conducted to determine the relationship between participants' concept-question scores and comprehension-question scores on the 40 Level 2 through Level 6 passages they collectively read (see Appendix B). For this analysis, only Level 2 through Level 6 passages were used because these passages all included four concept questions so the total possible score on this section was the same for all passages. The median, average, and range of PPVT-4 vocabulary sets administered to establish basal and ceiling scores were calculated. The words on the *Words Their Way ESI* (Bear et al., 2012) were ordered according to both word frequency in English and developmental spelling order and compared to ELs' spelling performance on the first administration of the ESI with a trendline analysis. Item 10 was excluded in the analysis because two participants took an earlier version of the ESI (Bear et al., 2007), which used the word *throat* rather than the word *shopping* for item 10. All other words were the same in the two versions. Qualitative analysis focused on the extent to which lan-

guage demands of the assessments may have influenced differences between ELs and pilot and normative samples.

Findings

Vocabulary Knowledge May Play a Role in ELs' Demonstration of Content Knowledge on the QRI-5

There is routinely a relationship between how much prior knowledge a reader has on the topic of a text and how well the reader comprehends that text. The QRI concept questions are intended to measure a reader's prior knowledge so that that information can be used to better understand how well the reader does on the comprehension measures after reading. For the ELs in this study, concept questions were somewhat predictive of passage comprehension, but vocabulary knowledge may also have played a role in ELs' ability to answer these questions. A Spearman rank analysis showed a moderate correlation between concept question scores and comprehension question scores on the 40 Level 2 to Level 6 passages the participants read ($r_s(40) = .41$, $p = .0087$). Although it is difficult to make a direct comparison between Leslie and Caldwell's (2006, 2011) pilot data and the findings in this study, because this study determined overall correlation between concept questions and comprehension questions for Levels 2 through 6 passages whereas Leslie and Caldwell provide overall correlations for upper middle-school and high-school passages and data about correlation on a few individual elementary-level passages, the overall moderate correlation between concept questions and comprehension questions in this study was similar to some of the correlations Leslie and Caldwell (2011) report for younger readers reading elementary passages. It was not as strong as the correlations the authors report for middle and high school students reading the middle- and high-school passages.

One possible explanation as to why the relationship between concept knowledge and comprehension was not stronger for the adolescent EL participants is that the passages analyzed were elementary-level passages. But another possibility is that concept questions that focused on the meanings of specific words and phrases may have assessed vocabulary knowledge as much or more than conceptual knowledge. Over a third of the 116 concept questions across the 50 passages participants read asked about the meaning of specific words or phrases. As mentioned above, this is a definitional free-association style (DFA), in which students are prompted to define a term (e.g., "What is X?") or identify a person (e.g., "Who was X?"). For individuals with only one language, there is a higher likelihood that the vocabulary that person has reflects the concepts he or she knows. ELs, on the other hand, could know a concept but know the word for it in another language. In such cases, the DFA style of question may not accurately capture concept knowledge.

The possibility that the concept questions assessed ELs' vocabulary knowledge was also noticed in the qualitative data. For example, when I asked two ELs, both talented artists, the concept question: *What is an illustrator?* neither participant could answer the question. One participant said: "Well, it's a noun. …Well, like something we use, or some person?" She identified the part of speech of the word but not the meaning. Similarly, when I asked an EL who was a member of the track and field team the question: *Why do people run races?* she was unable to give any answer, resulting in a score of 0. If a set of concept questions used a vocabulary term repeatedly, it had the potential to greatly influence an EL's score on the concept questions. For one passage, the word *beaver* appeared in all concept questions, beginning with the question *What is a beaver?* Neither of the ELs who read this passage answered any of the four concept questions, but it is unclear if they lacked prior knowledge about all the concepts or were unable to display their knowledge (e.g., knowledge of dams in the question *What are dams built by beavers?*) because they did not know the first word/concept *beaver*.

When I explained vocabulary that came up in concept questions that students could not answer, this sometimes helped ELs demonstrate prior knowledge. For example, when I told one EL what *wool* and *yarn* were, in the questions *What is wool used for?* and *What is yarn used for?*, after the participant had been unable to answer these questions, he answered the question *Why do people get haircuts?* by stating "make the clothes." Although this answer is incorrect for the question asked, it suggests that the EL did indeed have some knowledge about the topic of shearing sheep for wool.

Some concept questions featured polysemous words that led ELs astray. For example, to the question *What do waves do?* one EL responded "wave at your friend." In order to understand that this question is asking about waves in water, the examinee may need to have a sense of the likelihood that each type of wave could occur in plural form and together in a question with the word *do*. Answering the question *What does it mean to be gifted?*" another EL replied "give someone a gift … gift is gifted." In this question, the examinee would need to know that when *gifted* appears with the words *to be* it is more likely that it is referring to the meaning of having a talent than receiving a gift. Both of the answers ELs gave here received a score of 0 as "unconnected responses" (Leslie & Caldwell, 2011, p. 51) because the definition of the words that the participants gave were not the definitions for those words that would have indicated prior knowledge of the content of the passage.

In most cases, the concept questions had relatively simple syntax, but one Level 2 passage had questions with complex syntax. For example, the question *What can children do to keep themselves busy on long car rides in the car?* has three verbs (*should*, *do,* and *keep*), two prepositional phrases (*on long car rides*, *in the car*), and uses a reflexive pronoun (*themselves*). The EL who read this passage was unable to give an answer to this question.

Adolescent ELs May Take Longer to Complete the PPVT-4 *Because They Have Vocabulary Gaps*

The ELs who took the PPVT-4 in this study needed to respond to almost twice as many vocabulary sets as the fluent English speakers included in the normative sample (Dunn & Dunn, 2007) in order to complete the assessment. For the EL participants, the median number of vocabulary sets administered to establish a basal set (only one word wrong) and ceiling set (eight or more words wrong) was 9.5, whereas the normative sample required approximately 5 (Dunn & Dunn (2007). The range for ELs was 5-16 sets. For the nine participants who were administered both versions of the PPVT-4, the average number of sets for version A was 7.8, and the average number of sets for version B was 10.7. Table 2 indicates the version of the PPVT-4 that was used for each administration and the number of sets each participant completed. In one case, after progressing through 11 sets, the participant began losing concentration and did not complete the assessment.[2]

Table 2
Total PPVT-4 Vocabulary Sets to Establish Basal and Ceiling Scores

Participant	*English proficiency on MEPA*	*Time 1, number of sets*	*Time 2, number of sets*	*Average sets for participant*
Participant 1	1	6 (form A)	16 (form B)	11
Participant 2	1	5 (form A)	11 (form B)	8
Participant 3	2	11 (form A)	12 (form B)	11.5
Participant 4	3	7 (form A)	11+[a] (form B)	9
Participant 5	3	10 (form B)	10 (form A)	10
Participant 6	3	10 (form A)	10 (form B)	10
Participant 7[b]	4	9 (form B)		9
Participant 8[b]	4	7 (form A)		7
Participant 9	5	9 (form A)	7 (form B)	8
Participant 10	5	8 (form B)	7 (form A)	7.5
Participant 11	5	11 (form B)	5 (form A)	8

[a]Participant became fatigued and test was discontinued before ceiling was established.
[b]Participants 7 and 8 stopped participating in the study before the second administration of the PPVT-4, and therefore results for the Time 2 administration cannot be reported.

Interestingly, participants with greater English proficiency, on average, required a smaller number of PPVT-4 vocabulary sets than participants

with less English proficiency. The average number of sets for the three ELs with English proficiency scores of 1 or 2 was 10.2. The average number of sets for the three ELs with English proficiency scores of 3 was 9.7, and the average number of sets for the three ELs with English proficiency scores of 4 was 7.8. Two participants (see Table 2) were not included in this calculation because they stopped participating in the study before both versions of the PPVT-4 could be administered.

The ELs as a group also seemed more likely to know some words than others. Qualitative analysis indicated that within vocabulary sets, there were vocabulary items that were frequently answered correctly by ELs while other vocabulary items were infrequently answered correctly. For example, all 10 ELs who were administered set 9 of Form B (the recommended start set for age 10) correctly identified the picture of the word *angle*, but only three ELs identified the correct picture for *antlers*. Similarly, all 10 ELs who were administered set 11 of Form B (the recommended start set for age 13) correctly identified the picture for the word *triplets*, but only two identified the correct picture for *hoof*.

Vocabulary gaps could explain these findings, as gaps may have caused participants, especially those with lower English proficiency, to answer a few questions wrong on a wider range of vocabulary sets. Since the ELs in this study had all arrived in the US after age 10, they may not have been exposed to certain vocabulary that native English speakers usually acquire at a young age and, at the same time, may have acquired some grade-level academic vocabulary from their time in English-speaking schools. Vocabulary gaps could explain why ELs more frequently identified *angle* and *triplets* than *hoof* and *antlers*. Older ELs, such as the participants in this study, may be more likely to encounter the word *triple* or the word part *tri-* in high school courses than the word *hoof,* which might appear in storybooks. Another possibility is that some participants identified *angle* and *triplets* as cognates in their L1, although this could not explain Chinese, Japanese, or Korean participants' correct answers on these items.

ELs May Be More Likely to Spell Words That Occur Frequently in English Correctly on the ESI

The ESI is designed to measure developmental spelling level, so this analysis looked at how much developmental spelling level seemed to predict the words ELs spelled correctly, but it also considered the extent to which word frequency predicted correct spelling. Interestingly, developmental word order and word frequency appear to be related to how many words the EL participants in this study spelled correctly on the ESI. Figure 1 shows a trend line for developmental spelling according to the ESI and number of participants who spelled the word correctly, and Figure 2 shows a trend line for word frequency according to the *Corpus of Contemporary American*

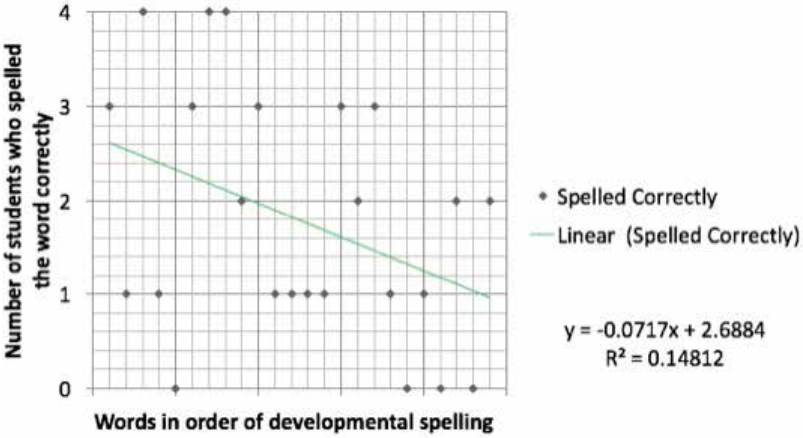

Figure 1. Trend line for developmental spelling order.

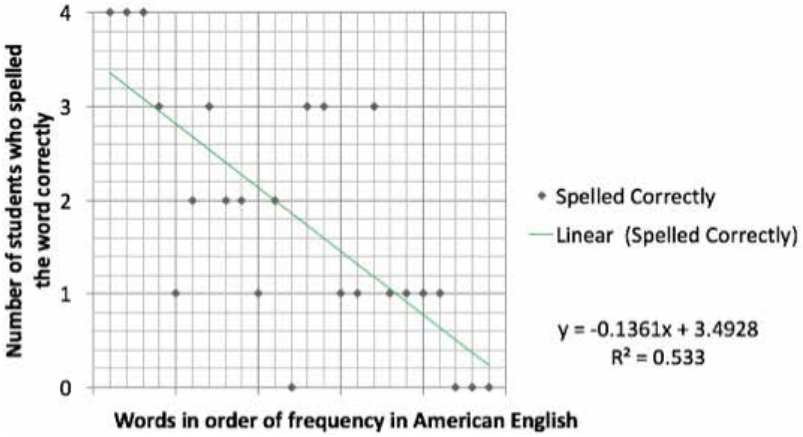

Figure 2. Trend line for word frequency.

English (Davies, 2008) and number of participants who spelled the word correctly. Developmental spelling order had a small goodness of fit proportion (R^2 = .15), and word frequency had a medium goodness of fit proportion (R^2 = .53). This finding seemed to differ from the item-difficulty indices from Sterbinsky (2007), which suggested that developmental spelling level was a better predictor of correct spelling than word frequency for the normative sample (which included ELs but did not look at ELs separately from all students).

A close look revealed that the three words all four participants in this study spelled correctly on the ESI were also the three words on the ESI that occur most frequently in American English, according to the *Corpus*

of Contemporary American English, and of the four words no participant spelled correctly, three words occurred the least frequently, compared to other words on the ESI, in American English (see Tables 3 and 4). Bear et al. (2007) indicate that inconsistencies in ELs' spellings may occur because ELs memorize the spelling of words beyond their developmental level. If the ELs in this study had memorized high-frequency words that were in their own English vocabulary, it could explain why word frequency appeared to be more strongly related to spelling than the developmental spelling level of words. ELs' spellings of English words may also have been influenced by orthographic and/or phonological transfer from their first languages, as other research has suggested (Martin, 2017). For example, one EL, who was literate in French, appears to have used her knowledge of French orthography when prompted to spell *favor*. She wrote *faveur,* which is the French cognate for *favor*. Additionally, ELs' first-language phonological systems may have influenced how they heard and spelled sounds. For example, the Korean participant and the Bulgarian participant wrote *sheep* for *ship*, which might be explained by the absence of a contrast between /i/ and /ɪ/ in their first languages (Klagstad, 1958; Lee, 2001). The Korean participant scored consistently low on final consonants, scoring only three out of five feature points for this phonic feature. Final consonants are usually mastered in the *emergent* (first) stage of spelling, so this result (a) was inconsistent with her power score—the total number of words she spelled correctly—which indicated that she was in the *syllables and affixes* (fourth) stage of spelling and (b) could be evidence of first language transfer, as Korean phonology allows fewer final consonants than English.

Implications

This exploratory study examined 11 ELs' performance on the QRI-5 and PPVT-4 and four ELs' performance on the ESI. The assessments gathered useful information about participants' literacy skills in English. At the same time, results suggest that the participants' performance on all three assessments may differ from fluent English speakers' because of linguistic demands. Although there was a relationship between scores on QRI-5 concept questions and comprehension questions for the ELs in the study, the data suggested that ELs may not have been able to thoroughly demonstrate their prior knowledge because they did not know the English vocabulary in some concept questions. ELs also took more time to complete the PPVT-4 because they needed to answer more vocabulary sets in order to get almost all words right on a basal set and at least eight wrong on a ceiling set. On the ESI, ELs' spelling appeared to be influenced by developmental spelling order but also by how common a word was in English and the extent to which words had phonological and orthographic features similar to participants' first languages.

Table 3
Words in Developmental Order with Number of Participants Who Spelled Each Correctly

Bed	3	Train	3	spoil	3	Shower	3	pleasure	1
Ship	1	Place	4	serving	1	Bottle	2	fortunate	0
When	4	Drive	4	chewed	1	Favor	3	confident	2
Lump	1	bright	2	carries	1	Ripen	1	civilize	0
Float	0	shopping	[a]	marched	1	Cellar	0	opposition	2

[a]Two participants took an earlier version of the ESI, which includes the word *throat* in place of the word *shopping*. For this reason, results were not calculated for this item.

Table 4
Words in Order of Frequency, as Established by the *Corpus of Contemporary American English* (Davies, 2008), With Number of Participants Who Spelled Each Correctly

When	4	bright	2	shopping	[a]	Serving	1	marched	1
Drive	4	Train	3	confident	2	Lump	1	Ripen	1
Place	4	Bottle	2	float	0	Spoil	3	cellar	0
Bed	3	opposition	2	shower	3	Chewed	1	fortunate	0
Ship	1	pleasure	1	favor	3	Carries	1	civilize	0

[a]Two participants took an earlier version of the ESI, which includes the word *throat* in place of the word *shopping*. For this reason, results were not calculated for this item.

In practice, some adjustments to these assessments may reduce the linguistic challenges discussed above and possibly increase the accuracy of the assessments. As Leslie and Caldwell (2011) note, some children in their pilot study who read pre-primer passages demonstrated concepts rather than explaining them verbally, leading the researchers to ponder whether they were "measuring verbal expression to a stranger rather than prior knowledge of specific concepts" (p. 455). Flexibility in *how* examinees may answer concept questions may support both ELs and native English speakers—in particular, younger readers. Also, linguistic modification has been found to be an effective and valid accommodation for ELs (Abedi & Gándara, 2006) and may mitigate the influence of language demands on ELs' performance. For example, rephrasing concept questions with difficult syntax on the QRI-5 could give ELs the opportunity to show more of what they know. The question above, *What can children do to keep themselves busy on long car rides in*

the car?, could be asked in a series of simpler sentences that repeat key ideas (e.g., *Sometimes children take a long trip in a car. They might get bored. What are some fun things children can do during a long trip in a car, so that they don't get bored?*). Teachers and test makers could also explore linguistic modifications to the ESI to ensure that test items are high-frequency words. Knowing the meaning of words in oral language provides an anchor for learning the written version (Helman & Burns, 2008). Furthermore, ELs' first-language backgrounds could be considered in developing test items and/or analyzing results, as certain types of sounds or spellings may be particularly challenging for examinees, depending on their first-language backgrounds.

Depending on the purpose for assessment, another appropriate accommodation could be first-language support during testing in English or additional assessment in ELs' home languages. Since even bilinguals with high levels of proficiency in both languages may know words for a given concept in only one of their languages (García et al., 2006a), QRI-5 concept questions may more accurately capture prior knowledge if they are translated or if ELs are given the opportunity to respond in their first languages. Concept questions do not influence miscue or comprehension scores that are used to determine instructional reading level, so this accommodation should not influence the outcome of the assessment, but it could potentially reveal more about what background knowledge an EL brings to the passage. At the same time, it is important to keep in mind that translating an instrument into another language is a complex task that will affect the psychometric properties of the assessment. Because of the design of the PPVT, translating test items would not be recommended, since vocabulary items may differ in frequency and therefore difficulty in different languages. Furthermore, since bilinguals tend to have different vocabulary in each of their languages, testing in only one language, even if it is the home or dominant language, may result in underestimation of their vocabulary knowledge (Kester & Pena, 2002). For example, Dunn (1988) found that bilingual Spanish-English speakers in the US did not perform as well as monolingual Spanish speakers on a Spanish version of the PPVT, the Test de Vocabulario en Imágenes Peabody: Adaptción Hispanoamericana (Dunn & Dunn, 1986).

Extended time is a common accommodation for ELs on assessments, yet making ELs focus on test items for long periods can lead to fatigue and potentially even affect results. Teachers can mentally prepare students for the length of assessments and/or break assessments into multiple sessions if they know that ELs are likely to take longer to complete the assessment than non-ELs. This type of adjustment may support ELs on the PPVT-4.

Finally, when interpreting assessment results, teachers can examine qualitative data alongside test scores, as was done in this analysis, and/or use a dynamic assessment approach to better understand what ELs know and can do. For example, on the ESI, analysis of phonic features can reveal

inconsistencies that are not evident in the power score alone. Additionally, Haitana, Pitama, and Rucklidge (2010) found that a dynamic-assessment approach helped reveal why certain items may have been answered incorrectly by examinees whose cultural backgrounds were not represented in the normative sample of the PPVT, and Clark and Kamhi (2014) suggest that a discussion approach rather than fixed questions may reveal more about examinees' prior knowledge on the QRI. Such an approach could also help expose ELs' interpretations of concept questions.

Although it was beyond the scope of this study to investigate cultural and experiential bias, this could potentially factor in ELs' performance on the three assessments and deserves attention in future research. Because concept questions on the QRI-5 are designed to assess prior knowledge and do not factor in overall reading scores, it would not be problematic if ELs had no knowledge of concept-question topics, but cultural and experiential differences between the examiner and the examinee could influence the extent to which an answer is considered complete or appropriate. Furthermore, cultural differences in word frequency (García et al., 2006a) as well as concepts and/or images (Haitana et al., 2010) may present bias when the PPVT is used with ELs. Development of the PPVT has included bias and sensitivity-review panels and analysis of item bias, but the perspectives represented (African American, Hispanic, Asian, Native American) may not have included the cultural perspectives of all EL students. On the ESI, experiential knowledge could also play a role in the words ELs spell correctly, and some of the sentences included to help examinees recognize spelling items on the ESI could potentially be more helpful to some individuals than others, depending on experience. For example, the sentence *It was fortunate that the driver had snow tires* for the test item *fortunate* may not provide a meaningful context for students who are unfamiliar with snow tires.

It is important to acknowledge that this analysis looked at assessment results of a very small number of ELs, and the participants were not selected as a representative group. Therefore, the results of this study cannot be generalized to other ELs.

The findings of this exploratory study, though limited in the ways noted above, illustrate how ELs may perform differently from native English speakers on the QRI-5, PPVT-4, and ESI because of the linguistic demands of the assessments. These demands include syntax, phonology, orthography, and especially vocabulary. The process also shows how teachers can examine assessments that they are using closely in order to (a) better understand and interpret ELs' results and (b) potentially make adjustments in the administration of these assessments when they are used informally to gather information about ELs' abilities. Without field testing with the EL population, these assessments should not be used to evaluate whether an EL's performance is typical or average compared to the performance of the groups

that were used in developing the assessments, but rather only as one source of information about an EL's English literacy skills, alongside other measures and qualitative analysis when possible. Yet, each can serve as a valuable assessment tool for teachers who understand how ELs' performance may differ from native and/or proficient English speakers'.

Author

Melissa Latham Keh, EdD, is an assistant professor in the TESOL Department at Bridgewater State University in Bridgewater, Massachusetts. She teaches courses on assessment and second-language literacy development. She formerly taught high school English and ESL for 12 years. Her research has focused on bilingualism across the lifespan and second language reading.

Notes

[1]*ELL, SPED,* and *gifted* are the terms used in Sterbinsky (2007).

[2]For purposes of analysis, this administration was included as 11 sets.

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment, 8*(3), 231-257. doi: 10.1207/S15326977EA0803_02

Abedi, J., & Gándara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice, 25*(4), 36-46. doi: 10.1111/j.1745-3992.2006.00077.x

Arias, G., & Friberg, J. (2017). Bilingual language assessment: Contemporary versus recommended practice in American schools. *Language, Speech, and Hearing Services in Schools, 48*, 1-15.

Bear, D. R., Helman, L., Templeton, S., Invernizzi, M. R., & Johnston, F. (2007). *Words their way with English learners: Word study for phonics, vocabulary, and spelling instruction* (1st ed.). Upper Saddle River, NJ: Pearson.

Bear, D., Invernizzi, M. R., Templeton, S., & Johnston, F. (2012). *Words their way: Word study for phonics, vocabulary, and spelling instruction* (5th ed.). Upper Saddle River, NJ: Pearson.

Clark, M. K., & Kamhi, A. G. (2014). Influence of prior knowledge and interest on fourth- and fifth-grade passage comprehension on the Qualitative Reading Inventory-4. *Language, Speech, & Hearing Services in Schools, 45*, 291-301. doi 10.1044/2014_LSHSS-13-0074

Davies, M. (n.d.). *The corpus of contemporary American English (COCA): 520 million words, 1990-present.* Available online at http://corpus.byu.edu/coca/

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assess-

ment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*(2), 145-186.

Dunn, L. (1988). *Bilingual Hispanic children on the U.S. mainland. A review of research on their cognitive, linguistic, and scholastic development.* Honolulu, HI: Dunn Educational Services.

Dunn, L. M, & Dunn, D M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). San Antonio, TX: Pearson Education.

Dunn, L. M., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody: Adaptción Hispanoamericana*. San Antonio, TX: Pearson Education.

Fernandez, M. C., Pearson, B. Z., Umbel, V. M., Oller, D. K., & Molinet-Molina, M. (1992). Bilingual receptive vocabulary in Hispanic preschool children. *Hispanic Journal of Behavioral Sciences, 14*(2), 268-276.

García, G. E., McKoon, G., & August, D. (2006a). Language and literacy assessment of language-minority students. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 597-624). Mahwah, NJ: Lawrence Erlbaum.

García, G. E., McKoon, G., & August, D. (2006b). Synthesis: Language and literacy assessment. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language minority children and youth* (pp. 583-596). Mahwah, NJ: Lawrence Erlbaum.

Goodman, Y. M., & Marek, A. (1996). *Retrospective miscue analysis: Revaluing readers and reading.* Katonah, NY: Richard C. Owens.

Goodman, Y. M., Martens, P., & Flurkey, A. D. (2014). *The essential RMA: A window into readers' thinking.* Katonah, NY: Richard C. Owens.

Haitana, T., Pitama, S., & Rucklidge, J. J. (2010). Cultural biases in the Peabody Picture Vocabulary Test-III: Testing Tamariki in a New Zealand sample. *New Zealand Journal of Psychology, 39*(3), 24-34.

Helman, L. A. (2005). Using literacy assessment results to improve teaching for English-language learners. *The Reading Teacher, 58*(7), 668-677. doi: 10.1598/RT.58.7.7

Helman, L. A., & Burns, M. K. (2008). What does oral language have to do with it? Helping young English-language learners acquire a sight word vocabulary. *The Reading Teacher, 62*(1), 14-19. doi: 10.1598/RT.62.1.2.

Jiménez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, *31*(1), 90-112.

Kester, E. S., & Pena, E. D. (2002). *Limitations of current language testing practices for bilinguals. ERIC Digest* (No. ED470203). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Klagstad, H. L. (1958). The phonemic system of colloquial standard Bulgar-

ian. *The Slavic and East European Journal, 2*(1), 42-54.
Latham Keh, M. (2014). *Second language reading of adolescent ELLs: A study of response to retrospective miscue analysis, error coding methodology and transfer of L1 decoding skills in L2 reading* (Unpublished doctoral dissertation). Boston University, Boston, MA.
Latham Keh, M. (2017). Understanding and evaluating ELLs' oral reading miscues. *Journal of Adolescent and Adult Literacy, 60*(6), 643-653. doi: 10.1002/jaal.625
Lee, J. (2001). Korean speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed.; pp. 325-342). Cambridge, England: Cambridge University Press.
Leslie, L., & Caldwell, J. S. (2006). *Qualitative reading inventory* (4th ed.). Boston, MA: Allyn & Bacon.
Leslie, L., & Caldwell, J. S. (2011). *Qualitative reading inventory* (5th ed.). Boston, MA: Allyn & Bacon.
Leslie, L., & Cooper, J. (1993). Assessing the predictive validity of prior-knowledge assessment. In D. J. Leu & C. K. Kinzer (Eds), *Examining central issues in literacy research, theory and practice* (pp. 93-100). Chicago, IL: National Reading Conference.
Martin, K. I. (2017). The impact of L1 writing system on ELL knowledge of vowel and consonant spellings. *Reading and Writing, 30*, 279-298. doi: 10.1007/s11145-016-9673-5
Massachusetts Department of Elementary and Secondary Education. (2010). *Overview of the Massachusetts English Language Assessment-Oral (MELA-O).* Retrieved from http://archives.lib.state.ma.us/bitstream/handle/2452/113443/ocn753988573.pdf?sequence=1
Nilsson, N. L. (2008). A critical analysis of eight informal reading inventories. *The Reading Teacher, 61*(7), 526-536. doi: 10.1598/RT.61.7.2
Peregoy, S., & Boyle, O. (2016). *Reading, writing, and learning in ESL: A resource book for teaching K-12 English learners* (7th ed.). Boston, MA: Pearson.
Sattler, J. M., & Altes, L. M. (1984). Performance of bilingual and monolingual Hispanic children on the Peabody Picture Vocabulary Test-Revised and the McCarthy Perceptual Performance Scale. *Psychology in the Schools, 21*(3), 313-316.
Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*, 209-231.
Sterbinsky, A. (2007). *Words their way spelling inventories: Reliability and validity analyses.* Memphis, TN: University of Memphis Center for Research in Education Policy.
Umberger, F. G. (1985). Review of the Peabody Picture Vocabulary Test-Revised. *TEST Critiques, 3*, 488-495.

Appendix A
Scoring for QRI-5 (Leslie & Caldwell, 2011) Concept Questions

3 points	"A Precise Definition, or a Definitional Response to a Phrase, or an Answer to a Question Specifically Related to Passage Content ... A Synonym" (p. 49)
2 points	"An Example of the Concept … A Specific Attribute or Defining Characteristics … A Function" (pp. 49-50)
1 point	"A General Association … Isolation of Prefix, Suffix, or Root Word … Firsthand, Personal Associations" (p. 50)
0 points	"Sound-Alikes … Unconnected Responses … No Response or I Don't Know" (p. 51)

Appendix B
QRI-5 Passages (Leslie & Caldwell, 2011) Read by Participants

Passage	***Level***	***Particpant***										
		1	2	3	4	5	6	7	8	9	10	11
The Brain and the Five Senses	1	X										
The Family's First Trip	2	X										
The Trip to the Zoo	3	X										
Cats: Lions and Tigers in Your House		X	X	X	X							
Wool: From Sheep to You			X	X								
The Friend			X			X						
Where Do People Live				X	X							
A New Friend from Europe					X							
A Special Birthday for Rosa					X	X	X					
Amelia Earhart	4		X	X		X						
The Busy Beaver				X	X							
Early Railroads				X		X						
Johnny Appleseed				X				X				
Tomie dePaola					X		X	X				
Plant Structures for Survival						X		X				
The Octopus	5			X	X		X					
Margaret Mead					X				X			
The Early Life of Lois Lowry	6			X								
Temperature and Humidity									X			
Abraham Lincoln									X			
The Lifeline of the Nile									X			
Pele											X	
Life Cycles of Stars—1	UM								X			
Biddy Mason										X	X	
Malcolm X										X		
Where the Ashes Are—Part 1	HS									X	X	X
Where the Ashes Are—Part 2											X	
World War I—Part 1											X	
Characteristics of Viruses—Part 1												X