Pearson TELL

New Technologies for Assessing Language Learners

By Paula Hidalgo, Ph.D.; Edynn Sato, Ph.D.; and Masanori Suzuki

More than 10 percent of students in the United States are English language learners (ELL) – a number that has doubled over the past 15 years. In fact, it is projected that by 2025 that one out of every four students will be an English language learner. While 70 percent of all ELL students live in Arizona, California, Texas, New York, and Florida, ELL student populations have grown over 200 percent in the last 10 years in Tennessee, North Carolina, South Carolina, Georgia, Indiana, and Nevada.

At the same time that the ELL population is experiencing this rapid growth, schools are faced with the challenge of ensuring that all students – including ELLs – meet today's new rigorous learning goals and are prepared for success in college and careers.

However, a performance gap is evident in annual statewide summative assessments, and the need to continue to support ELLs' English proficiency development is apparent in states' Annual Measurable Achievement Objectives (AMAO). Each state develops and implements AMAOs for holding all Title III-funded local education. The gap between ELLs and non- ELLs, at least among adolescents, is affected by students' literacy skills, which both depend on and promote their knowledge of academic language (Alvarez, Ananda, Walqui, Sato, & Rabinowitz, 2014; Snow & Biancarosa, 2003). Consequently, reducing the performance gap necessitates a focus on language development and literacy (Alvarez et al, 2014). Currently, limited resources are available to teachers and students that provide information in a timely, formative manner about students' language development and progress toward learning goals so instruction can be adjusted to meet students' dynamic learning needs.

This paper presents theory- and research-based considerations behind the design and development of Pearson's TELL™ (Test of English Language Learning) and its performance-based tasks. TELL is a formative assessment tool intended to support students' development of English language proficiency.

More specifically, the paper focuses discussion on contextualization and authentic uses of language, skill integration, and progress-monitoring to support English language proficiency development. It is followed by a description of TELL and the pilot study results that informed its development.

Contextualization and Authentic Uses of Language

Research suggests that "deep learning" is facilitated through contextualization (Moltz, 2014). Such contextualized learning involves the presentation of basic skills within topics or situations that can allow for meaningful, authentic application of skills, as well as the linking of ideas and concepts in a manner that promotes learning and transfer (Alvarez et al, 2014; Echevarria, Vogt, & Short, 2012; Heller & Greenleaf, 2007; Howard, Sugarman, & Coburn, 2006; Lee & Spratley, 2010; Perin, 2011; Simpson, Hynd, Nist, & Burrell, 1997). When knowledge and skills related to English language proficiency are not contextualized, the learning of targeted knowledge and skills generally has been shown to be less effective than when they are contextualized (Grubb, 1999; Perin, 2011; Simpson et al, 1997). Contextualization of language knowledge and skills includes global tasks or functions, such as asking for information, describing events, and expressing opinions, as well

as situations, such as communicating in school or in a restaurant, so that language is used or applied in a meaningful way (American Council on the Teaching of Foreign Languages, 2012; TESOL, 2006). Pearson TELL presents elements of language in familiar and meaningful ways to facilitate language use and learning.

Integrated Skills

Language learning is a dynamic, developmental process, and research supports an integrated approach to the teaching and learning of English, which generally reflects how children learn language (Walsh Dolan, 1985; Levine & McCloskey, 2013; Li, 2012). Rather than being addressed separately, the modalities of listening, speaking, reading, and writing, as well as related language skills and register, for example, are best taught and learned in an integrated manner (Levine & McCloskey, 2013; Li, 2012). For example, oral English has been shown to facilitate English reading; therefore, engaging students' listening and speaking skills supports students' reading ability development (August & Shanahan, 2006; Short & Fitzsimmons, 2007). Consistent with research on language learning, TELL integrates skills, as appropriate (e.g., reading and writing, reading and speaking, listening and speaking), to promote language proficiency development. In addition, most real-life language use situations require multiple language skills. For example, in conversations, one has to listen to the other person's speaking and then respond to it appropriately in speaking. Another example is that in the classroom, one has to read written texts and then summarize main points with some supporting details in writing. Integration of language skills in TELL's assessment tasks, therefore, ensures authentic use of language skills in real life.

Progress Monitoring

The capacity to monitor students' progress in learning – involving a continuous cycle of gathering evidence and evaluating student learning, providing feedback to students about their learning, and using data to adjust subsequent instruction as needed

– is essential in facilitating student achievement (Alvarez et al, 2014; Black & Wiliam, 1998; Heritage, 2011). This capacity is particularly valuable as it can pertain to the dynamic nature of English language proficiency development.

Therefore, in progress monitoring, as students engage in tasks, evidence of their current learning is gathered and evaluated to yield immediate feedback specifically linked to what the student is trying to learn or accomplish, and the feedback leads to subsequent thinking and action by the student that is intended to further his or her learning. Such engagement has value in placing the student at the center of learning, promoting student agency in learning (e.g., for self-regulated learning, self-efficacy) and embeds the assessment within an instructional activity so instructional time is optimized (Hattie & Timperley, 2007; Heritage, 2008; Marshall & Drummond, 2006).

TELL's primary use is progress monitoring to support and promote instruction as described above. In addition, progress-monitoring tests are coupled with a screener and diagnostic assessments because data are needed at different points in a learner's development to effectively facilitate learning. The need for comprehensive data as much as possible is critical in language proficiency development because of its dynamic developmental nature.

Pearson TELL: A Tablet-Based Assessment Tool for K – 12 English Language Learners

With consideration of theory and research, Pearson developed TELL, a tablet-based assessment tool to support the development of English language proficiency of K-12 ELLs. TELL consists of tasks that can be used for screening, diagnostic, and progress-monitoring purposes.

- The screener test is used to determine if a student qualifies as an English language learner and therefore is entitled to benefit from available programs for ELLs.
- The diagnostic tests (pre- and post-) are used at the beginning of the year to establish a baseline,

and at the end of the year to compare and provide measures of growth during that period. The diagnostic tests are the most robust and the most thorough of all three test types.

· The progress-monitoring tests are sets of eight forms for grades K-12 that can be used monthly to track the growth of the language skills at regular intervals. TELL divides grades K-12 into five sub-grade bands – K, 1-2, 3-5, 6-9, and 9-12 – to provide tasks and content suitable for particular age ranges.

In total, TELL features 21 item types, as summarized in Table 1 below. Many item types integrate multiple language skills to elicit student performances that typically provide information about multiple language skills and practical language features. Most test forms also include items that probe specific foundational skills, which are required to meet some state standards. TELL scores report information about a student's English abilities in listening, speaking, reading, and writing as well as other subskills such as grammar, vocabulary, fluency, pronunciation, and pre-literacy.

Table 1: Summary of TELL Item Types

Item Type	Item Despcription	Skills	Grade Band
Say the word	Verbally identify image shown on screen	Listen-Speak	K
Pick the right picture	Identify by touch, the desired basic text feature from three images	Read/Print literacy	K
Say the letter	Read list of 5 upper- and lower-case letters aloud	Write/Print literacy	K
Copy the letter	Copy the letter displayed on the screen	Write/Print literacy	K
Copy the word	Copy the word displayed on the screen	Write/Print literacy	K, 1–2
Pick the right letter (for K)	From a group of 3 letters or 3 words, touch the one that represents the sound played	Read/Print literacy	K, 1–2
Pick the right letter (for 1–2)			
Find the error	Identify the word with a spelling or capitalization error	Read/Print concepts	1–2
Write the word	Handwrite the word that is heard with corresponding image	Write	1–2
Write about the picture	Write a description for the picture shown on screen	Write	1–2
Read the words	Read list of words aloud	Read-Speak	1–2
Describe the video	View silent video and describe its events	Speak	1–2
Listen and act	Follow audio orders to touch, drag, move or draw a path	Listen	K, 1–2, 3–5, 6–8, 9–12
Repeat the sentence	Listen to a short audio and repeat verbatim	Listen-Speak	K, 1–2, 3–5, 6–8, 9–12
Listen and retell	Listen to an audio passage and retell the story	Listen-Speak	K, 1–2, 3–5, 6–8, 9–12
Read and act	Read prompt then touch, drag and drop an object	Read	1–2, 3–5, 6–8, 9–12

Item Type	Item Despcription	Skills	Grade Band
Complete the sentence	Read a sentence with missing word and then type the word in the blank	Read-Write	1–2, 3–5, 6–8, 9–12
Put the word in the sentence	Correct grammer of scrambled sentence by drag and drop	Read	3–5, 6–8, 9–12
Speak in the situation	Respond to an audio and graphic prompt	Listen-Speak	3–5, 6–8, 9–12
Watch and explain	Watch a video of a teacher explaining a concept then answer a comprehension question	Listen-Speak	3–5, 6–8, 9–12
Read the passage	Read text aloud and answer comprehension question verbally	Read-Speak	3–5, 6–8, 9–12
Read and summarize	Read a text passage, answer a comprehension question, and then write a summary	Read-Write	3–5, 6–8, 9–12

Table 2 shows an example for a diagnostic test for grade levels 3-5. The table shows 11 item types, with mention of the number of items corresponding to each type. This sample comprises a total of 42 items with an estimated test time of 50 minutes. The last col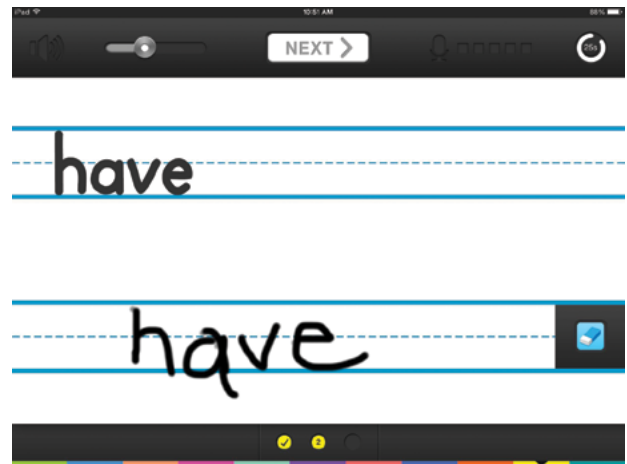umn points to the main modalities of each item, indicating whether the item is discrete or integrated. The sample test structure was tested in January, February, and March 2015. The table does not illustrate the test structure of the final product to be released in summer 2015.

Table 2: Diagnostic Test Design

Section	Number of Items	Task-Item Type Level: Grade 3, 4, 5	Main Modalities
Introduction			
Type 1	4	Listen and act	Listen
Type 2	8	Repeat the sentence	Listen-Speak
Type 3	4	Speak in the situation	Listen-Speak
Type 4	3	Listen and retell	Listen-Speak
Type 5	1	Watch and explain	Listen-Speak
Type 6	2	Read the passage	Read-Speak
Type 7	4	Read and act	Read
Type 8	6	Put the word in the sentence	Read
Type 9	8	Complete the sentence	Read-Write
Type 10	2	Read and summarize	Read-Write
Closing			
Total	42		

Act by Speaking

In this item, the student is given an audio prompt that describes a situation with a question at the end that the student is expected to answer orally. In this case, the audio prompt is: "You have a group project due soon. Members of your group are talking about meeting at the library Saturday afternoon. You remember you already have plans with your family all day Saturday, but you are available Saturday night. What do you say to your group?" The screenshot includes the microphone (which is active), the three green audio dots (which indicate that the student is speaking at the right noise level), the timer (counting down), and the next button (which is available to tap when the student is finished speaking).
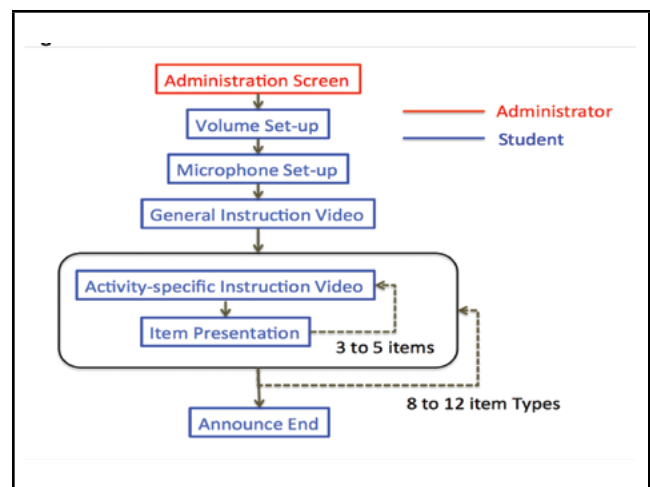

Copy the Word

Copy the Word is a K-2 item type only. The student is given an audio prompt saying, "This is the word x. Write the word x here." In this case, it is the word have. The student then uses his or her finger to write the word on the second set of lines at the bottom. If the student makes a mistake, he or she can tap the eraser icon, which erases everything. Once the eraser icon is tapped, it changes to an undo icon, which brings back the writing, if tapped. The Next Button is active and the timer is counting down.

Presentation of the Test

Below is a visual representation of how the test is presented. By identifying the role of the administrator/teacher/proctor (in red) and the role of the student (in blue), it is clear to the reader that most of the steps are performed by the student, thereby highlighting one of TELL's most important features, namely, that it does not require proctoring, and consequently instructional time is fully protected. This is one of the key differentiators of TELL, and can result in a great source of savings for districts, schools, and teachers.

Figure 1: Presentation Flow

Pilot Studies

To examine the effectiveness of TELL's tasks, two pilot studies were conducted in 2012 (Bernstein et al, 2013). The first one was conducted with 326 students, and the second was conducted with 458 students (Total N=784). The following were the research questions addressed:

· Which activities/tasks/items work?
· Which yield the most information?
· Which discriminate ELLs from other students?

Various test-like sequences of items were presented to the children, ages 4 to 11. Fifty-three percent were from non-English-speaking homes, and 47 percent were in official ELL status at the time the pilots were conducted. The sequences selected for the pilots comprised between 24 and 45 items. These sequences were designed to cover many combinations of input and output modalities that are available through a tablet. Some of the activities were designed to elicit information about a single skill (e.g., either listening or writing), while others were designed to elicit information on several integrated skills (e.g., both listening and writing or reading and writing). The first types of activities are referred to as discrete and the second type as integrated. All activities were tested on iPads. Materials were presented in several modalities (e.g., speech only, speech with picture, text), and responses were obtained in several modalities (e.g., speech, typing, drawing, dragging).

Procedures

Groups of three to eight students ages 4 to 11 were given an iPad in the same room doing the same activities, but not in synchrony with each other. After the administrator entered the ID and selected the appropriate test form, students were left to go through the test on their own, guided by video tutorials. A general video tutorial was presented at the beginning of the test to help students with general user aspects (e.g., how to go to the next item). Short, specific videos on each item type preceded groups of four to five items presented for each type to help students understand how to perform that particular task. Most children ages 4-7 encountered 36 items, while those students ages 8-11 encountered approximately 43 items.

Results

Generally, the results of pilot tests to date support the importance of contextualized, meaningful tasks, as well as the viability of presenting language modalities and/or language skills and features in an integrated manner. Additional studies will be conducted to further examine the effectiveness of TELL's tasks in providing timely, formative information about students' language development and progress toward learning goals – so instruction can be adjusted to meet students' dynamic learning needs, as well as to inform refinement of TELL's tasks.

Automated Scoring Technology in TELL

Unlike conventional language assessments, a unique feature of TELL is that student responses are automatically scored including spoken and written responses. The automated scoring systems implemented in TELL are custom-built for TELL items. However, TELL uses technology that has been researched, validated and operationalized for more than a decade as part of Pearson's automated language assessments and language instructional systems.

Over the last 15 years, Pearson has scored millions of spoken and written constructed responses for use in primary, secondary and post-secondary education, as well as by governments, publishers and corporations from all over the world.

The underlying technology for evaluating spoken responses is based on the patented Versant testing system, which uses a speech processing system optimized for speech from native and non-native speakers of the language tested. In addition to recognizing words, the system locates and evaluates relevant segments, syllables and phrases. The Versant testing system then uses statistical modeling techniques to assess the spoken performance of the test-taker. Currently, the Versant scoring

technologies have successfully been applied for Arabic, Chinese, Dutch, English, French and Spanish.

Pearson's technology for automatically scoring written responses uses Latent Semantic Analysis (LSA), a statistical language learning theory and computer model that measures the semantic similarity of words and documents with accuracy closely approximating that of human judges.

This automated written response scoring technology is used in highstakes testing contexts such as the writing section of Pearson Test of English – Academic and classroom products for building literacy skills, such as WriteToLearn and Summary Street.

Accuracy of Automated Scoring

Independent studies have shown that Pearson's automated scoring systems are objective and produce measures of reliability that exceed those of many standard human-rated tests.

Table 3 below presents correlation coefficients between scores produced by the automated scoring systems and scores provided by expert human raters for a sample of written and spoken constructed responses. Column 4 (Machine-Human Score Correlation) shows the correlation between automatically-generated scores and human rater-based scores. Column 5 (Human-Human Score

Correlation), on the other hand, presents correlation coefficients between two human raters. For example, the third row shows score accuracy indicators for a set of five informationintegration items. For each item, students were asked to write memos that synthesized information from multiple sources, including letters, memos, summaries of research reports, newspaper articles, maps, photographs, diagrams, tables, charts and interview notes or transcripts.

A set of 1,239 written responses were produced and scored by the automated scoring system as well as by independent raters. The observed correlation coefficient in the "Machine-Human" column suggests that machine scores reasonably match a good human score derived by combining the human scores to form a stable average score. This correlation exceeds the correlation of 0.79 between pairs of scores given by pairs of individual skilled human raters, shown in the rightmost

"Human-Human" column. For these memos, automatic scores are closer to a stable consensus human score than one expert score is to another.

Results from other validation studies for both spoken and written responses in the table demonstrate similar patterns, indicating that the machine scores are very similar to expert human scores.

Table 3: Sample sorrelation coefficients between machine scores and human scores

Autoscoring Performance				
Response	Assessment Prompt Material	n	Machine to Human Score Correlation	Human-Human Score Correlation
Written	81 published essay prompts (grade 6–12)	400	0.89	0.86
	18 research-leveled essay prompts (grade 4–12)	635	0,91	0.91
	5 information synthesis memorandums using a range of information sources	1239	0.88	0.79
Spoken	2000 spoken English tests taken by adults, diverse items types	143	0.97	0.99
	3000 spoken Arabic (diverse item types)	134	0.97	0.99
	9 Oral Reading Fluency passages for 1st – 5th grade	248	0.98	0.99

Summary

Available for back-to-school 2015, TELL combines a strong research foundation with the power of technology to provide a cutting-edge solution for assessing English language development progress in a fun and engaging tablet environment.

References

Alvarez, L., Ananda, S., Walqui, A., Sato, E., & Rabinowitz, S. (2014). *Focusing formative assessment on the needs of English language learners.* San Francisco: WestEd.

American Council on the Teaching of Foreign Languages (2012). *ACTFL Proficiency Guidelines 2012.* Alexandria, VA: American Council on the Teaching of Foreign Languages.

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

Bernstein, J., Todic, O., Newmeyer, K., Schulz, K., & Zhao, L. (2013). *Young Children's Performance on Self-administered iPad Language Activities.* Menlo Park, CA: Pearson.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*(1), 7–73.

California Department of Education (2014a). *Facts About English Learners in California—CalEdFacts.* Retrieved from http://www.cde.ca.gov/ds/sd/cb/cefelfacts.asp.

California Department of Education (2014b). *Glossary of Terms.* Retrieved from http://www.cde.ca.gov/ds/sd/cb/glossary.asp#u.

Echevarria, J., Vogt, M.E., & Short, D.J. (2012) *Making Content Comprehensible for English Learners: The SIOP Model* (4th Edition). New York, NY: Pearson.

Grubb, W. N. (1999). *Honored but invisible: An inside look at teaching in community colleges.* New York, NY: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Heller, R., & Greenleaf, C. L. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement.* Washington, DC: Alliance for Excellent Education.

Heritage, M. (2008). Learning progressions: *Supporting instruction and formative assessment.* Washington, DC: Council of Chief State School Officers.

Heritage, M. (2011, Spring). *Formative assessment: An enabler of learning. Better: Evidence-based education.* Retrieved from http://www.cse.ucla.edu/products/misc/bettermagazineheritage.pdf

Hill, L. E. (2012). *California's English Learner Students.* Retrieved from http://www.ppic.org/main/publication_quick.asp?i=1031.

Howard, E. R., Sugarman, J., & Coburn, C. (2006). *Adapting the Sheltered Instruction Observation Protocol (SIOP) for Two-Way Immersion Education: An Introduction to the TWIOP.* Washington D.C.: Center for Applied Linguistics.

Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy.* New York, NY: Carnegie Corporation of New York's Council on Advancing Adolescent Literacy.

Levine, L. N., & McCloskey, M. L. (2013). *Teaching English language and content in mainstream classes: One class, many paths,* 2nd ed. Boston, MA: Pearson.

Li, J. (2012). *Principles of Effective English Language Learner Pedagogy.* College Board.

Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education, 21*(2), 133–149.

Molle, D. (2015). *Academic Language and Academic Literacies: Mapping a Relationship.* In Molle, D., Sato, E., Boals, T., & Hedgspeth, C. (Eds.) Multilingual Learners and Academic Literacies: Sociocultural Contexts of Literacy Development in Adolescents. New York, NY: Routledge.

Moltz, D. (2014). *Encouraging deep learning.* Inside Higher Ed. Retrieved from http://www.insidehighered.com.

Perin, D. (2011). F*acilitating Student Learning Through Contextualization.* New York, NY: Community College Research Center, Teachers College, Columbia University.

Short, D., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners.* Washington, DC: Alliance for Excellent Education.

Simpson, M. L., Hynd, C. R., Nist, S. L., & Burrell, K. I. (1997). College academic assistance programs and practices. *Educational Psychology Review, 9*(1), 39–87.

Snow C. E., & Biancarosa, G. (2003). *Adolescent literacy and the achievement gap: What do we know and where do we go from here?* New York: Carnegie Corporation of New York.

Teaching English to Speakers of Other Languages (2006). *Pre-K–12 English Language Proficiency Standards.* Alexandria, VA: TESOL.

Walsh Dolan, M. (1985) "Integrating Listening, Speaking, Reading, and Writing in the Classroom," *Language Arts Journal of Michigan:* Vol. 1: Iss. 1, Article 4.

To explore TELL, visit **PearsonAssessments.com/TELL**

800-328-5999 | PearsonAssessments.com

Pearson