Assessing Language performance with the Tablet English Language Learner Assessment (TELL)

Paula Hidalgo
Pearson Knowledge Technologies

Key Words: Language, automated & tablet

Abstract

New technologies in touch-tablet computing enable integrated combinations of language activities (listening, speaking, reading, watching, writing, manipulating) in engaging environments for students. This paper presents the development and piloting of a formative assessment with 25 different English language activities implemented on a touch-tablet (iPad) device. The performance tasks integrate combinations of the four language skills with nonlinguistic skills. For example, oral narration while watching a video, listening and moving objects around the screen, and summarizing texts in writing.

The assessment uses automated speech and writing scoring and is intended for use with preK-12 students whose first language or home language is not English. The assessment was trialed with 700 English language learners and native English students. Learners' performances were captured to provide diagnostic proficiency estimates to language standards.

The paper discusses item design and development, demonstrate items, and present pilot results, including an analysis of the kinds of automatic language measurement that these items can support. The paper further discusses how information about student abilities can be used to:

1.   inform English language program placement decisions
2.   profile student abilities and skill levels to inform instruction
3.   monitor progress to improve formative assessment information and predict performance on summative tests

Paula Hidalgo

paula.hidalgo@pearson.com

International Association for Educational Assessment (IAEA) 2014 Conference

Submission Number: **120**

Presentation Title: *Assessing Language performance through the Tablet English Language Learner Assessment (TELL)*

Full Paper

1. The state of English Language Learning

English language learning is growing rapidly, both in countries where the home language is English as well as countries in which English as a second language has become a critical requirement. In the United States, bilingualism has been on the rise, with the number of school-age children who speak a language other than English at home estimated at around 11.2 million (over 20%). Of these students, approximately 5.5 million US students are enrolled in some form of English language learner (ELL) program, whether English as a Second Language, High Intensity Language Training, bilingual education, dual immersion, etc. And these numbers continue to increase. Over the past 15 years, the number of ELLs in these programs has doubled. By 2015, the ELL enrollment is expected to grow to 10 million. By 2025, 1 out of every 4 U.S. student is expected to be an English Language Learner (National Education Association, 2008).

Globally, data from the Ambient Insight Premium Report show that the global market for digital English language learning products and services reached $1.31 billion in 2011 and continues to grow rapidly, and it is estimated that revenues will reach 2.6 billion by 2016. It is estimated that over 1 billion people are currently learning English worldwide. According to the British council, as of the year 2,000 there were 750 million of people learning English as a foreign language (EFL) and 375 million learning English as a second language (ESL). (English as a Foreign Language speakers use English occasionally for business or pleasure, while English as a Second Language speakers use English on a daily basis).

In both the US and the majority of other countries, current methods of assessments tend to be traditional, relying on paper and pencil, multiple choice, and/or drill and practice.   These methods limit students' ability to learn language in engaging interactive contexts and receive

rapid feedback.   In addition, instructors seldom have the resources to monitor and help all students in a timely fashion. For example, to practice and receive feedback on speaking skills, students must often speak in front of a teacher, although a teacher can only assess one student at a time.   With the growth in need for practice, this approach is not sustainable.

In this presentation we discuss and demonstrate a prototype of an assessment currently in development by Pearson, designed to provide an engaging environment in which students interact with a tablet (in this case an iPad) much like a game, while performing activities that allow assessment of a broad range of language skills. The TELL assessment uses the Versant automated scoring for spoken and written responses, thereby automating much of was heretofore done by instructors. This assessment is self-delivered, thereby allowing teachers to maximize instructional time and minimize administration and proctoring time.

2. Pilot testing of the TELL

This assessment was pilot tested during the years 2012 and 2013, with over 700 low SES children. 53% of them came from non-English speaking homes. The pilot yielded 28,000 responses to self-administered activities (items) which were modeled by a video example.

The goal of the pilot was to measure K-6 language skills in contextualized, authentic, real-life situations.   Researchers used self-administered tasks modeled with a short video in a fully automatic way (administration, presentation, data capture, upload & scoring) to elicit performances that provide rich information on language skills.

Analysis of the students' responses was designed to answer three questions:
   a. Which activities (items) can children understand well enough (without adult help) to perform meaningfully?
   b. Which specific activities yield the most information about a child's relative language skills, and which materials are appropriate for ages 4-7 and 8-11?
   c. Which activities best discriminate English language learners (ELLs) from 'mainstream' students?

Analysis of the more than 28,000 responses show that most of the activities (assessment items) can be successfully modeled by a single short instructional video example, included in the test, and most of them elicit useful responses from students. The few activities that did not elicit useful information are being eliminated or replaced for the final launch of the assessment. By age 7 or 8, nearly all children respond meaningfully to about 95% of these specific activities, regardless of first language.

3. The TELL Assessment

TELL is a combination product and service that includes three elements:

   • a touch tablet delivery platform, e.g. iPad or Android tablet, with software to present & score TELL items

- a pool of automatically scored items sufficient to assemble the various test forms described below. Automatic scoring has been used very successfully to assess written, spoken, and to a lesser extent mathematical responses. Written and spoken automated scoring have been combined to assess the traditional four language skills (reading, writing, speaking, and listening) for college admission and employment decisions. (For most aspects of writing and speaking, the performance of automated scoring already equals or surpasses that of human raters).
- a back-end system configured to store responses and scores, along with an API or UI for score access.

TELL test forms are intended for use with students in PreK through grade 12, whose first language or home language is not English. TELL scores will report information about a student's abilities in English listening, speaking, reading, and/or writing in relation to "referenced standards", that is, to the standards in force for local users.

Depending on the TELL assessment form, test scores can be used for three main purposes:

Screener/Placer: to inform program placement decisions for English language development

Diagnostic Profiler: to profile the abilities and skill levels in order to inform instruction

Progress Monitor: to monitor student progress.

The Diagnostic Profiler test will have the longest test form among the three. The Screener/Placer and the Progress Monitor tests are comprised of a subset of item types included in the Diagnostic Profiler test.

Elements of the test construct can be organized into the following categories:

Modalities: listening, speaking, reading, writing

Register: social, general academic (e.g. "tier 2" vocabulary), discipline-specific (e.g. math, social studies, earth science)

Skills: spelling, print awareness, fluency, pronunciation, oral reading, grammar, vocabulary, usage, anaphora, conventions (punctuation, capitalization, etc.), meta-linguistics (language used to describe language)

Practical Language Features: organization, inference, pragmatics.

TELL forms will feature integrated performance tasks that combine multiple receptive modalities to elicit student performances that typically provide information about ability in multiple language skills and practical language features. Most test forms will also include some items that probe specific foundational skills, which are required to meet referenced standards. Many of the integrated-skill task types and specific-skill item types support the development of

items that will probe a wide range of cognitive complexities and levels of referenced standards. Item developers will write items that meet specific cells of referenced standards using the item types described below. These item types are also appropriate for learners in international markets, where item content can be developed to meet international standards and to predict performance on established international tests.

TELL tests are fixed forms that present from 20 to 60 items, and are designed for self-administration on a touch tablet computer. Depending on the form and its purpose, these items are organized into 1 to 11 sections. Each section presents an instructional video specific to the task-item type and then 1 to 8 task-items of that one type. The order of items within an item-type section may be partially random. The figure below presents the test administration and item presentation flow.
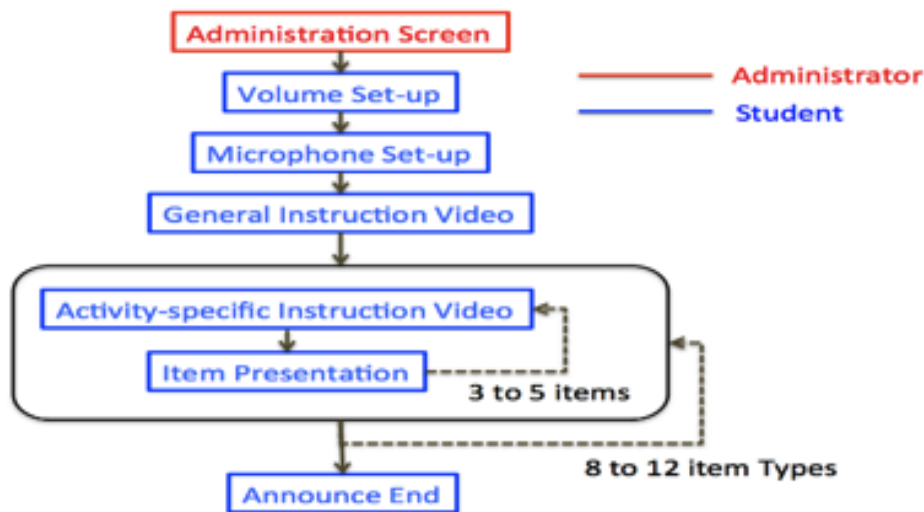

Figure 1: Test administration and item presentation flow.

TELL test forms will be designed within a 3 x 6 schema: Test Purpose (3): Screener/Placer, Diagnostic Profiler and Progress Monitor) and Test Levels (6): (PreK, K, 1-2, 3-5, 6-8, 9-12)**.**

The TELL schema is shown in the table below. Test length varies according to test use and level, as indicated in the table. The Diagnostic Profiler test is always the longest. Both the Screener/Placer and Progress Monitor tests are expected to have the same structure. Item types for these two tests are selected from the corresponding grade-level Diagnostic Profiler test. Furthermore, the Progress Monitor test will not be administered every week. It is expected to be administered at a frequency of every 4 to 6 weeks.

The following table presents an example design for a 36-minute TELL Diagnostic Profiler that presents 40 items within 11 task-item types, and is intended for students in grades 3–5.

Section	No. of Items	Task-Item Type Level: Grade 3,4,5	Time in Minutes	Main Modalities
Introduction			1	
Type 1	4	Listen & Act	1	Listen
Type 2	6	Listen & Repeat	1	Listen-Speak
Type 3	4	Act by Speaking	1	Listen-Speak
Type 4	2	Listen, Answer, Retell	4	Listen-Speak
Type 5	2	Listen, Watch, Explain	4	Listen-Speak
Type 6	4	Read & Act	2	Read
Type 7	5	Order Sentence	2	Read
Type 8	2	Read Aloud & Answer	4	Read-Speak
Type 9	5	Complete Sentence	3	Read-Write
Type 10	4	Use Modifiers	2	Read-Write
Type 11	2	Read, Answer & Summarize	10	Read-Write
Closing			1	
TOTAL	**40**		**36**	

Table 6. Example design for a 36-minute Diagnostic Profiler test with 40 items within 11 task-item types for grades 3–5.

Note that the example above proceeds from oral language to text and from receptive to productive modalities. An implementation of this example test form should elicit language-enabled responses from students in grades 3–5 that can be analyzed by computer to produce scores for the principal skills and practical language features, across modalities in multiple registers.

Schools could promote a holistic development through the use of this 4-skill assessment by integrating it with their curricula. The assessment contains automated speech and writing scoring capabilities. Written texts are scored for declarative knowledge and language skills as reflected in stylistic and mechanical aspects of writing. Spoken responses are scored for declarative knowledge and speech quality in tasks such as reading aloud to determine fluency and in orally summarizing a reading.

4. Demonstration

During the presentation, participants will have the opportunity to see a short demo (2-3 minutes long) of four assessment items. The presenter will project the items using her iPad and will perform the role of the student to show participants exactly how the assessment is used (self-administered) and how the system reacts when a student performs a task. The items the presenter will demonstrate include:

1. <u>Oral Read</u>: This item type presents a short paragraph that is read aloud by the system. Students are asked to read aloud after they hear the voice-over. They then have to orally answer a comprehension question about what they read.

2. <u>Read Move</u>: Students are presented with a short command accompanied by a picture. They read the command (they can read it in low voice or aloud) and they perform the action requested, which generally entails moving objects with their finger.

3. <u>Silent Video</u>: This item type shows a short silent video (about 5 seconds long). The student is asked to orally explain what is happening in the video.

4. <u>Choose the word</u>: Students are presented with a short sentence/paragraph accompanied by an illustration. Some of the words are highlighted in color. Students are orally instructed to tap on the highlighted words and change the word if it makes the story better. When students tap the word, two additional choices are presented. Students tap the word of their choice which is automatically moved to the specific section in the sentence where it belongs. Below is a screen-shot of an item of this category.

5. **Conclusion**:

TELL is designed as an engaging assessment that allows students to show what they know while 'playing' and while responding to fun, real-life activities on a device they love. Most activities feel like a game and students tend to forget they are being assessed.

Both hard and soft technologies (Shute & Zapata-Rivera, 2008), as well as educational, emotional and psychological measurement approaches have advanced significantly in the past couple of decades. It is now possible to accurately and efficiently diagnose student competencies at various levels during the course of learning. With TELL, we can now accurately and effectively measure and automatically score language performance using stealth assessment (i.e., data are collected while students play with their ipads on language activities) and allowing teachers more time to focus on instruction instead of administering the test.

Bibliography

Ambient Insight Premium report. The Worldwide Market for Digital English Language Learning Products and Services: 2011-2016 Forecast and Analysis, The Emergence of a Global Technology-based English Language Learning Ecosystem.

NEA (National Education Association). English Language Learners Face Unique Challenges. Policy Brief, 2008.

Shute, V. J & Zapata-Rivera, D. (2008). Adaptive technologies. In J. M. Spector, D. Merrill, J. van Merriënbook, & M. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (3rd Edition) (pp. 277–294). New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

Streeter, L., Bernstein, J., Foltz, P., & Deland, D. (2011). Pearson's Automated Scoring of Reading, Writing and Mathematics. Retrieved from
http://www.pearsonassessments.com/research