

# TELPAS Listening and Speaking Cognitive Lab Summary Report

## Background

The Texas English Language Proficiency Assessment System (TELPAS) measures the progress that English language learners (ELLs) make in acquiring the English language. TELPAS consists of annual statewide English language proficiency assessments for ELLs in grades K–12 in the language domains of listening, speaking, reading, and writing. Currently, the TELPAS assessments are performance-based and holistically rated, with the exception of the reading assessments for grades 2–12, which are multiple-choice tests. For each language domain, TELPAS measures four levels, or stages, of increasing English language proficiency: beginning, intermediate, advanced, and advanced high.

Starting in 2017–2018, the listening and speaking domains for grades 2–12 will no longer be holistically rated. These holistic ratings will be replaced with an online assessment of both listening and speaking that will include multiple-choice items, technology enhanced items, and automated-scored speaking items. In 2015–2016, educator focus groups were convened, and item development and review began. In spring 2017, a statewide pilot study and field test will be conducted for the redesigned listening and speaking assessments.

## Purpose

In September 2016, a cognitive lab was conducted to obtain feedback from ELLs on the design of the new TELPAS listening and speaking assessments. Information on two design components was collected.

- A usability study component was included to observe whether students could access and navigate the online system without difficulty.
- A content study component was included to evaluate whether item types resulted in the types of listening and speaking responses they were designed to elicit.

These activities provided an initial qualitative look at whether the online layout allows students to navigate easily through each of the items, and from item to item, and whether students can engage with the content of the items to demonstrate their listening and speaking skills.

Information obtained from the cognitive lab will be used to refine the online interface and item types in preparation for the spring 2017 pilot.

## Study Design

### Participants

The study took place with volunteer districts in and around the Austin area. All four grade bands (2–3, 4–5, 6–8, and 9–12) were included. Since the cognitive lab took place in the fall, students participating were assigned the grade band from the prior school year. For example, grade 4 students participated in the 2–3 cognitive lab. Campuses were asked to include students from a variety of proficiency levels and grade levels, so that results will be applicable to beginning, intermediate, advanced, and advanced high students and all grade levels within a grade band. A target sample of 36 students was selected to be included in the cognitive lab, 8–12 students per grade band and 9 students per proficiency level.

*Table 1. Target Cognitive Lab Population*

Grade Band	Proficiency Level				Total
	Beginning	Intermediate	Advanced	Advanced High	
2–3	3	3	3	3	12
4–5	2	2	2	2	8
6–8	2	2	2	2	8
9–12	2	2	2	2	8
<b>Total</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>36</b>

The cognitive lab was administered to 36, but due to the limited availability of students on the days when observers were on campus, the exact target sample was not obtained. For example, only one advanced student participated in the grades 4–5 cognitive lab, while three beginning level students participated in the grades 6–8 cognitive lab. Every item set included in the cognitive lab was shown to at least three students at a variety of proficiency levels. Of the students taking part in the cognitive lab, there was a variety of first languages spoken, including the two most common first languages other than English spoken in Texas households (Spanish and Vietnamese).

### Items and Item Types

For the listening audio prompts and selected-response items, students listened to the audio prompt and responded to items on the computer. For the speaking prompts, students spoke their responses into a microphone that captured audio responses on the computer. There were 14 different item types included in the cognitive lab. Three of these item types were presented in multiple formats. This included listening items, speaking items, and integrated items that tested both listening and speaking. The plan was designed to evaluate a selection of item types at each grade band. Not all item types were presented during the cognitive lab. If multiple item types had similar functionality and layout, only one of those item types was included. Additionally, item types with functionality that was already familiar to students were not included in the cognitive lab. The item types presented in the cognitive lab by grade band are shown in Table 2.

For all grade bands except 2–3, each student responded to one of two item sets consisting of five items each. Three item sets, consisting of four items each, were used for the 2–3 grade band. Each student responded to both listening and speaking items in a single grade band. Items were selected based on a review of content and usability features to make sure a variety of items appeared during the cognitive lab.

*Table 2. Item Sets for Cognitive Lab*

<b>Grade Band</b>	<b>Set 1</b>	<b>Set 2</b>	<b>Set 3</b>
<b>2–3</b>	Picture Sequence (L) Ask a Question (S) Explain a Process-visual (S) Compare Pictures-chunked (S)	Picture Matching #3 (L) Speak in the Situation (S) Describe a Picture (S) Passage Based, part 2 (L/S)	Passage Based, part 1 (L) Explain a Process-prompt only (S) Compare Pictures-not chunked (S) Tell a Story (S)
<b>4–5</b>	Picture Matching #2 (L) Explain a Process-prompt only (S) Compare Pictures-not chunked (S) Open Ended (S) Passage Based, part 2 (L/S)	Passage Based, part 1 (L) Explain a Process-visual (S) Compare Pictures-chunked (S) Point-Counterpoint (L/S) Tell a Story (S)	
<b>6–8</b>	Picture Sequence (L) Passage Based, part 1 (L) Explain a Process-visual (S) Speak in the Situation (S) Ask a Question (S)	Picture Matching #2 (L) Passage Based, part 2 (L/S) Compare Pictures-chunked (S) Point-Counterpoint (L/S) Describe a Picture (S)	
<b>9–12</b>	Picture Matching #3 (L) Passage Based, part 2 (L/S) Open Ended (S) Describe a Picture (S) Tell a Story (S)	Passage Based, part 1 (L) Simulated Conversation (L/S) What do you Think (S) Compare Pictures-not chunked (S) Explain a Process-prompt only (S)	

### Procedures

In order to familiarize students with the online interface, TEA and Pearson staff guided students through a sample item before viewing the study items. As the students reviewed the study items, they were prompted to “think aloud” about their experience interacting with the online system and how easy or difficult it was to respond to each item type. Additionally, interviewers asked students a series of scripted questions following each item that were designed to evaluate the usability of the online system and content engagement (whether students interpreted and responded to the listening and speaking items as intended). An example of the interview protocol, which includes the scripted questions for Set 1 of grades 4–5, is presented in the Appendix.

For beginning and intermediate proficiency-level students, the individuals conducting the study communicated with the students in their first languages when needed. The interviews lasted between 20–45 minutes.

## Summary of Results

The cognitive lab study, including information from the student think-aloud, interviewer observations of the students, and student responses to interviewer questions, was designed to help answer the following questions:

- Do students understand the directions for each item type?
- Are the item layouts clear for students to navigate?
- How often do students access the audio for the speaking item stimuli (text that is also onscreen)?
- How often do the students rerecord their spoken responses? Are their responses well within the allowed time, or do they run out?
- Do the items we are hoping to write 4-point rubrics for result in extended responses?
- For the items presented in multiple formats, which format was more successful, and was it different by grade?
- Are there any items types that seem cognitively inappropriate for any grade band?

Do students understand the directions for each item type?

In general, students did not have any problems understanding the directions for each item type. In the few cases when students did have initial confusion about the items, after a re-reading of the “gray-box” directions (item directions that are static for all items of an item type), or by getting some clarification from the cognitive lab facilitator, the students were able to complete the items. All help of this type, whether it involves reading-aloud, translating, or explaining of the gray-box directions and not part of the item stem, would be allowable by a test administrator during an actual administration.

For some of the speaking item types, it was observed that students only produced the minimum required amount of speech to answer the items. Thus, for some item types, the speaking item stems will be changed to try and encourage the students to produce more speech.

Are the item layouts clear for students to navigate?

For nearly all of the item types, students did not have any trouble navigating the item layout or using the online interface. The one exception to this was for items that had multiple scrolling capabilities, one for the item and one for the passage. The presence of multiple scrolling options was confusing to some students. These students did not observe that they had multiple methods of scrolling, and would try to scroll the entire item when they needed to scroll the passage portion only. To minimize confusion, items of this type will be represented in the tutorial to make sure all students are aware of multiple scrolling capabilities.

How often do students access the audio for the speaking item stimuli (text that is also onscreen)?

Almost all of the students accessed the audio for speaking items, especially at the lower grade bands. Sometimes these students would only listen to the item prompts without reading them, and other times they would both listen to and read the prompts. Only students in the grades 9–12 grade band read the items without listening to the prompts for all of the items.

How often do the students rerecord their spoken responses? Are their responses well within the allowed time, or do they run out?

Students rarely rerecorded their audio responses. Even when listening to their response, most students were satisfied with keeping their initial response and not rerecording. All student responses were well within the allowed time, with no student running out of time while trying to record a response.

Do the items we are hoping to write 4-point rubrics for result in extended responses?

The speaking item types that were designed to elicit responses scored on a four-point rubric (one point for each proficiency level) produced responses that varied in length and proficiency. None of these item types were deleted due to a lack of student-produced speech. For some item types, however, the gray-box directions and some item stems will be slightly altered in an attempt to elicit more language (e.g., adding “at least 2 ways” or “as much as you can” to the end of the directions).

For the items presented in multiple formats, which format was more successful, and was it different by grade?

The only item types that were presented in competing formats were the “Compare Pictures-chunked” and “Compare Pictures-not chunked” item types. The intent of the “chunked” item type was to elicit multiple responses that could be separately scored, and the “not chunked” version asked the students to perform both (combined) speaking tasks for a single item score. Across all of the grade bands these “chunked” items did not elicit enough speech for multiple separate items. Therefore, the pilot test and any further item development will include only the “Compare Pictures-not chunked” item type.

Are there any items types that seem cognitively inappropriate for any grade band?

No items appeared to be cognitively inappropriate for students across proficiency levels in any of the grade bands.

## Conclusion and Next Steps

Overall, students had little to no difficulty navigating through or working with items. The few minor issues that were occasionally observed will be addressed either in the tutorial or by modifying the gray-box directions. Beginning- and intermediate-level students did have problems answering items, but the think-alouds and interviews provided evidence that this was due to limited language proficiency and not due to an inability to navigate through the items or understand what each task involved.

Gray-box directions and item stems will be slightly modified to encourage more speech before the pilot test in spring 2017. After the pilot testing, when student performance data can be analyzed, more information will be known about what item types will be available for the first operational administration of the redesigned TELPAS listening and speaking assessment in spring 2018.

## Appendix: TELPAS Listening and Speaking Cognitive Lab/Usability Study Interview Protocol

### **Instructions for Facilitators/Notetakers**

This study will be conducted using interview teams made up of at least one staff member each from TEA and Pearson. Each team will be observing and interviewing students for a given grade band. All members of the interview team will participate in the interview process through interactions with the students. Prior to each interview, one member should be designated the notetaker and the other member the primary interviewer. Teams may alternate roles if desired. Prior to each interview, Pearson staff will ensure that the computers and necessary hardware and software are working and ready.

### **Primary Interviewer Functions:**

- Introductions and instructions
- Answer student questions and provide clarification when needed
- Guide student through the sample item
- Prompt the student to think aloud if needed
- Ask follow-up questions about each item type as needed
- Take notes as feasible
- Debrief with team after each interview

### **Notetaker Functions:**

- Take detailed notes during interview
- Answer student questions and provide clarification when needed
- Prompt the student to think aloud if needed
- Ask follow-up questions about each item type as needed
- Debrief with team after each interview

Each team member should be familiar with the script that follows so the interview process is standardized to the degree possible. However, the script does not need to be read word-for-word to each student. The script should guide the interview team but should be tailored to the student's level of English proficiency or transadapted as needed. The provided follow-up questions should be tailored according to student observations and student feedback.

## **Introduction**

You are helping us with a very important task today. We are designing a new test to see how well students understand and speak English. We are here today to see if you understand how to use the computer to answer some questions. This is not a test. We are talking to some students that know a lot of English and some students that only know a little bit of English. We want to see if everyone can understand how to answer the questions on the computer.

We are going to show you some questions on the computer. We want you to look at them and talk to us about what you think you should do on the computer to answer the questions. We want to see if it is easy to figure out what buttons to push and if the buttons do what you think they will do. It is okay if you do not know the correct answer to the questions. Tell us whatever you are thinking about or have questions about. It helps us a lot if you are honest. Remember that this is not a test, and you are not going to get a grade.

While you are looking at the questions and talking about them, we will be taking some notes and asking you some more questions about what you are seeing and doing on the computer. This will help us create better questions.

Do you have any questions before we start?

If the student indicates that there is a question, answer it unless it is covered in the next section. When there are no further questions, load the form so the sample item screen is showing.

## **Functionality**

Let's look together at some things on the screen. At the top of every screen you will see some words in a gray box. <point to gray box> Those are instructions for that question. You can read those yourself. If you have trouble reading the instructions or do not understand them, I will help you.

Now we are going to look at some buttons together. When you see something that looks like this on the screen <point to audio play button>, it means there is something to listen to. Click on the button to hear some information. <wait for audio to play> You may listen to the audio again if you want by clicking on the button again. You can listen to it as many times as you want. Sometimes you will see this button under words on the screen like it is here. <point to the prompt> Then you can choose to listen to the information, read it, or both. Other times, you will see this button on the screen without any words to read. Do you understand how to use this button?

Make sure the student understands how to use the audio play button.  
Answer any additional questions before proceeding.

Now let's look at a different kind of button. When you see something that looks like this on the screen <point to speech capture button>, it means you will need to talk to answer a question and the computer will record what you say. Do you remember what you heard when you clicked on this button? <point to audio play button> If not, you can read or listen to it again now.

When you are ready to answer, you will click here <point to speech capture button> to start recording yourself. After you click on this button, start talking into this microphone. You need to make sure your mouth is near the microphone when you are talking so the computer will record you. When you are finished talking, click on the button again to turn the recording off. You will have two chances to record your answer. Let's try it. Are you ready to answer? If not, you can read or listen to the prompt again.

When the student is ready to answer the question, help the student launch the speech capture and turn it off. Prompt the student to say something if needed so there will be something to relisten to. The computer may prompt you at this point to identify the microphone in use. If you want to have the student wear the headphones for a moment to record this answer (and then remove them), go ahead and get the headphones on and sized, and the microphone placed. Otherwise, wait to do this until you turn control of the computer over to the student after modeling the think aloud and just have the student hold the microphone for now.

Now you have recorded yourself answering the question. After you record an answer, you can listen to what you said. Click on this button here to listen to the answer you recorded. <point to the play button to allow the student to play the file> If you like that answer, you can choose to keep it. If you do not like it, you can choose to delete that answer and try recording it one more time. You only get two chances to record your answer. If you want to get rid of your answer and try again, click here. <point to trash button> Would you like to keep your answer or delete your answer and try again?

If the student chooses to delete and try again, make sure the student clicks the correct buttons and listens to the retake before proceeding.

If you delete your answer and record a second answer, you can listen to it, but you cannot delete your second answer. You only get two chances to record your answer. Do you understand how to record yourself talking?

Make sure the student understands the functionality of the speech capture. Answer any additional questions before proceeding.

### **Modeling**

Now that you understand how to use the computer to answer the questions we are going to look at, I want to show you one more thing. We call this “thinking out loud.” The reason we call it that is because we want you to tell us out loud everything you are thinking about in your head. I am going to show you an example. <navigate back to sample screen> Ready? Just watch me and listen to what I am thinking about.

Start by modeling how to put on the headphones and adjusting the microphone.

Ok, first, I am going to put on these headphones and make sure they fit on my head so my ears are covered. Then I need to make sure the microphone is right in front of my mouth. Now I’m ready.

I’m looking at a computer screen. The first thing I see at the top is some words in a gray box. I am going to read those words. <model reading quietly> Ok, so now I understand that I am going to look at something and record myself speaking on this question. I look at the rest of the screen. Under the gray box, I see some words. Under the words I see that button we practiced that will tell me what the words say. I am going to click on that button so I can listen to the computer tell me what to do. <model clicking on audio play button> Wait, I don’t quite understand that so I’m going to listen one more time. Ok, so now I know that I understand what I need to do. <pause briefly to study art where applicable and talk about it> To answer this question, I want to remember to talk about [this is item specific; model preparing a response]. Now I’m ready to record my response. I click on the button with the red circle, and then begin talking. <click on speech capture button and record a response, then turn it off> Now I listen to my answer to see if I like it. I know that I need to click on the button with the triangle to listen. <listen to recorded response> I think that answer is ok, so I’m not going to delete it and record another, but I know that I could if I wanted to. I am finished with this question and ready to go to the next one.

Do you have any questions about what you heard me say or saw me do when I was answering this question?

Answer any final questions before preparing to give control of the computer to the student.

### **Instructions**

Now you are going to look at a group of questions. There are going to be different kinds of questions. Some might be easy for you to answer, and some might be more difficult. Remember that this isn't a test, and you aren't getting a score. We just want to see if you can figure out what you need to do on the computer to answer the questions. While you are looking at each question, tell me everything you are thinking about. You can also ask questions. Here is some scratch paper and a pencil you can use to take notes if you want. When you finish the first question, we will stop for a minute and I might ask you a question or two before we look at the next one.

Go ahead and put these headphones on. Are you ready to start?

- When the student is ready, go to the first cog lab item.
- The notetaker will record information on the notetaking document. The interviewer will take notes as feasible.
- After the student completes each item, make sure to ask the follow-up questions not already observed before moving on.
- Prompt the student as needed.

## Item-Specific Questions and Observations

GRADES 4-5, FORM B, SET 1

### Picture Matching

Did the student understand what to do on this item?	
Did the student understand how to use the drag-and drop functionality?	
Did the student listen to the audio multiple times?	
How many pictures did the student match correctly?	

**Open Ended (with art)**

Did the student understand what to do on this item?	
Did the student listen to or read the prompt?	
Did the student seem distracted by or overly dependent on the art in the response?	
Did the student understand how to use the speech capture functionality?	
Did the student listen to the response?	
Did the student delete the first response and record again?	
Did the student have sufficient time to respond? Was there significant time remaining?	
How would you rate the response to this item on a 4-point scale?	

**Compare the Pictures (not chunked)**

Did the student understand what to do on this item?	
Did the student listen to or read the prompt?	
Was the student able to successfully use the speech capture functionality?	
Did the student delete the first response and record again?	
Did the student address both parts of the prompt in the response?	
Did the student have sufficient time to respond? Was there significant time remaining?	
How would you rate the response to this item on a 4-point scale?	

### Open Ended

Did the student understand what to do on this item?	
Did the student listen to or read the prompt?	
Did the student delete the first response and record again?	
Did the student have sufficient time to respond? Was there significant time remaining?	
How would you rate the response to this item on a 4-point scale?	

The next item uses a video player. Most students probably know how to launch a video on the computer. However, if the student you are working with is not familiar with the functionality, use the language below to help.

On this screen you can see something that looks like a rectangle with a triangular play button on it. If you ever saw someone play a video or movie on a computer, it might look like that. If you see this, you will click on the arrow button to watch and listen to a short video. Go ahead and click on the play button now. [<point to the button to launch the video, then wait while the video is playing>](#)

**Passage Based (with speaking item)**

Did the student understand what to do on this set of items?	
How many times did the student listen to the complete video before proceeding to items?	
Did the student play the item-specific video chunks, or rely on memory from listening to the entire video?	
Did the student understand that the item-specific video chunks were different?	
Did the student listen to or read the items and prompt?	
Did the student correctly answer the MC item(s)?	
Did the student delete the first response and record again?	
Did the student have sufficient time to respond? Was there significant time remaining?	
How would you rate the response to this speaking item on a 4-point scale?	