

The ASSETS Consortium English Language Proficiency Assessment for Grades 1-12:

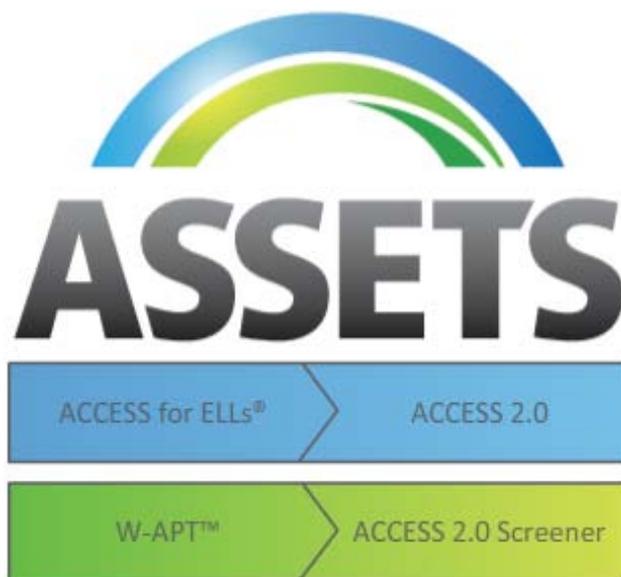
Test and Item Design Plan for the Annual Summative and On-demand Screener

Version 1.1 Update

April 19, 2013

Prepared by:

Center for Applied Linguistics



Center for Applied Linguistics
Washington, DC



Contents

1	Introduction	1
1.1	What is this document?	1
1.2	How to use this document	1
1.3	Summative test and screener	1
2	Background	3
2.1	Two consortia: WIDA and ASSETS	3
2.2	ASSETS assessment system	5
2.3	Levels of documentation	8
2.4	ACCESS and ACCESS 2.0	10
3	ACCESS 2.0 at a Glance	11
4	Test-level Concerns	14
4.1	Design and functionality	14
4.1.1	General principles	14
4.1.2	Design elements	15
4.1.3	Navigation	16
4.2	Accessibility and affect	19
4.2.1	Instructions	19
4.2.2	Virtual test administrator	20
4.2.3	Variation across grade levels and language proficiency levels	20
4.3	Key academic language functions	21
4.4	Adaptivity	21
4.4.1	Order of subtest administration	22
4.5	Tutorial	22
4.6	Paper-based accommodation	23
4.7	Reporting	25
4.7.1	Scores reported	25
4.7.2	Score reports	26
5	Subtest Specifics: Listening	27
5.1	Composition of the Listening subtest	27
5.2	Adaptivity	28
5.3	Item presentation	29
5.4	Domain-specific design elements	30
5.5	Student response	30
5.6	Scoring	31

5.7	Listening screener	31
6	Subtest Specifics: Reading.....	31
6.1	Composition of the Reading subtest.....	31
6.2	Adaptivity	32
6.3	Item presentation	32
6.4	Domain-specific design elements	33
6.5	Student response	33
6.6	Scoring.....	34
6.7	Reading screener.....	34
7	Subtest Specifics: Writing	35
7.1	Composition of the Writing subtest.....	35
7.2	Tier placement	35
7.3	Task presentation.....	36
7.4	Domain-specific design elements	36
7.5	Student response	37
7.6	Scoring.....	38
7.7	Writing screener	38
7.7.1	Composition of the Writing screener.....	38
7.7.2	Scoring the Writing screener	39
8	Subtest Specifics: Speaking	40
8.1	Composition of the Speaking subtest	40
8.2	Tier placement	40
8.3	Task presentation and student response.....	40
8.4	Domain-specific design elements	42
8.5	Scoring.....	43
8.6	Speaking screener	44
8.6.1	Composition of the Speaking screener	44
8.6.2	Scoring the Speaking screener	44
9	Next Steps	45
10	Appendix: Understanding the ACCESS for ELLs® Test	46
10.1	Introduction	46
10.2	Test structure.....	46
10.3	Tiers.....	47
10.4	Test administration times	49

1 Introduction

1.1 What is this document?

The *ASSETS Consortium English Language Proficiency Assessment for Grades 1-12: Test and Item Design Plan for the Annual Summative and On-demand Screener* builds on *The ASSETS Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener* to provide additional detail about (a) the new technology-delivered annual summative assessment of academic English language development, ACCESS 2.0, and (b) the on-demand ACCESS 2.0 Screener, a test given to incoming students who may be designated as English language learners.

Like the *Framework*, the *Test and Item Design Plan* is a living document that can and will evolve in response to feedback from stakeholders from throughout the ASSETS Consortium. Note that this document describes the **operational version** of the ACCESS 2.0 and ACCESS 2.0 Screener assessments; due to vendor specifications, some minor details of the new assessments described in this document may not be fully present in the **field test version** of the assessments.

1.2 How to use this document

The *Test and Item Design Plan* (TIDP) is complementary to the *Framework*. The TIDP includes more detail about some topics—the specifics of the domain subtests, for instance—than the *Framework*, and the *Framework* contains background information—about the WIDA ELD standards, for example—absent from the TIDP. To increase the TIDP’s ability to function as a stand-alone document, however, some sections from the *Framework* appear more or less verbatim in the TIDP. Readers familiar with the *Framework* may feel free to skip such sections. Readers who need more background on the current ACCESS for ELLs® assessment should consult the appendix, “Understanding the ACCESS for ELLs® Test,” at the end of the *Test and Item Design Plan*.

1.3 Summative test and screener

This document handles both (and only) the annual summative and screener components of the ASSETS assessment system (see section 2.2). The ACCESS 2.0 Screener can be thought of as a shorter version of ACCESS 2.0. The desire to keep the screener short, however, must be balanced with the need for it to yield a reliable measure of a test taker’s English language proficiency. More screener items will be administered to students functioning at the upper end of the proficiency scale because it is there that the decision of whether or not to designate a student as an English language learner is made.

Differences between ACCESS 2.0 and the ACCESS 2.0 Screener will be noted throughout this document as necessary and are summarized in Table 1.

Table 1. Differences between ACCESS 2.0 and the ACCESS 2.0 Screener

Aspect of test	ACCESS 2.0	ACCESS 2.0 Screener	Where to find more information
Purpose and Uses	<ul style="list-style-type: none"> secure, annual measure that meets federal requirements for monitoring ELLs' progress toward English language proficiency can serve as one of multiple measures used to determine student preparedness to exit English language support services 	<ul style="list-style-type: none"> on-demand, locally-scored measure that meets federal requirements as an instrument to help determine a student's ELL status in a timely fashion can serve as one of multiple measures to determine if a student is an English language learner qualified for English language support services 	Framework 3
Composition	<ul style="list-style-type: none"> Listening: 6-8 folders of test questions (17 to 26 questions) Reading: 8-10 folders (23 to 32 questions) Writing: 3 tasks, one of them extended (Tier B/C only) Speaking: 5 folders (10 tasks) 	<ul style="list-style-type: none"> Listening: 4-6 folders (11 to 20 questions) Reading: 4-6 folders (11 to 20 questions) Writing: 2 tasks, one of them extended Speaking: 1 task plus 3 folders (7 tasks) 	TIDP 5.1, 5.7, 6.1, 6.7, 7.1, 7.7, 8.1, 8.6
Scoring	<ul style="list-style-type: none"> Speaking responses centrally scored using a six-level rubric keybarded Writing responses or scanned handwritten responses centrally scored 	<ul style="list-style-type: none"> Speaking responses scored locally by educators trained using the ACCESS 2.0 Screener Rater Training Program—Speaking keybarded Writing responses or handwritten responses scored locally by educators trained using the ACCESS 2.0 Screener Rater Training Program—Writing 	TIDP 7.6, 7.7.2, 8.5, 8.6.2
Administration	<ul style="list-style-type: none"> test administrators need not be trained to score student responses for any domain 	<ul style="list-style-type: none"> test administrators of the paper-based version of the screener must be trained to score student speech for the Speaking domain during administration 	Framework 8

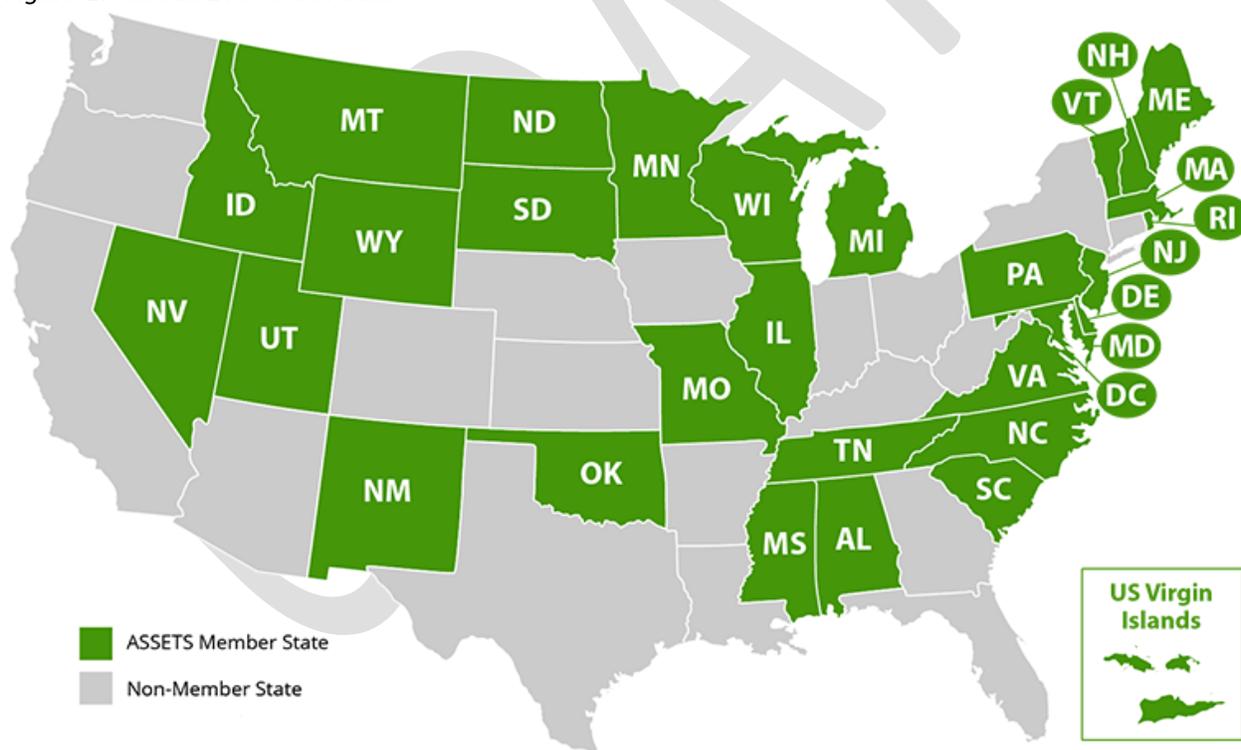
Reporting	<ul style="list-style-type: none"> • electronic data files and score reports will be produced in and distributed from a central location for distribution through state and district offices as appropriate 	<ul style="list-style-type: none"> • student Speaking and Writing scores are entered into an interface that combines them with the stored Listening and Reading scores, calculates the various composite scores, and generates electronic data files and score reports for local use 	TIDP 4.7
-----------	--	---	----------

2 Background

2.1 Two consortia: WIDA and ASSETS

The *Assessment Services Supporting ELs through Technology Systems (ASSETS)* Consortium was created through funding awarded by the U.S. Department of Education in September 2011 to the state of Wisconsin via an Enhanced Assessment Grant. Figure 1 shows the states that were members of the ASSETS Consortium as of April 15, 2013.

Figure 1. The ASSETS Consortium

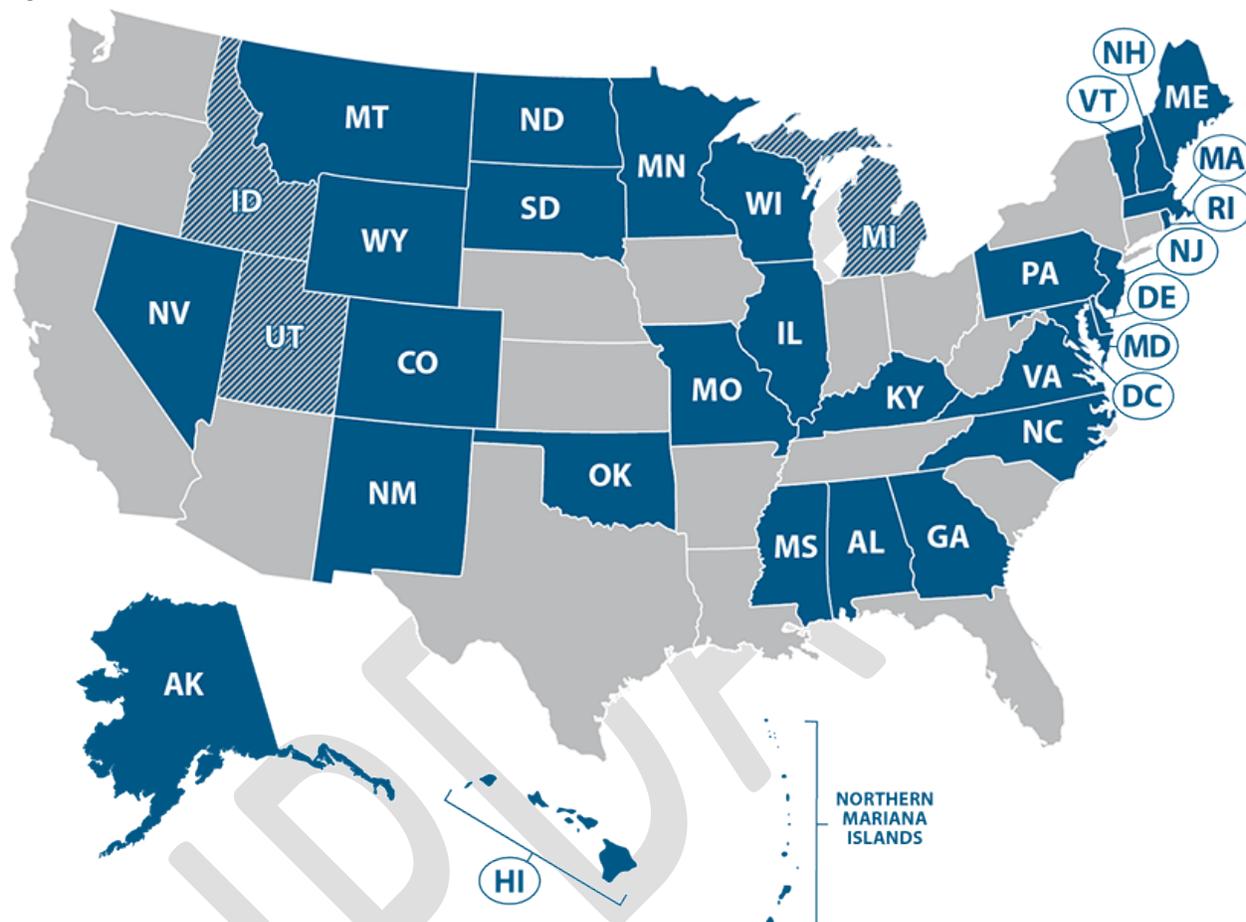


(<http://assetsproject.org/aboutus/memberStates.aspx>)

The proposal submitted by the ASSETS Consortium sought funding to build upon the earlier assessment efforts of the World-Class Instructional Design and Assessment (WIDA) Consortium. The WIDA Consortium, currently housed within the Wisconsin Center for Education Research at the University of

Wisconsin—Madison, was established in 2003, also with funding from a U.S. Department of Education Enhanced Assessment Grant awarded to the state of Wisconsin. As of April 15, 2013, the WIDA Consortium included 31 states. Figure 2 shows the states that are members of the WIDA Consortium.

Figure 2. The WIDA Consortium



States labeled and colored solid blue are members of the current WIDA Consortium. The three striped states have adopted the WIDA English Language Development Standards but do not participate in other WIDA Consortium activities. (<http://www.wida.us/membership/states/index.aspx>)

Since its beginning, the WIDA Consortium (hereafter WIDA) has created and expanded on comprehensive English language development (ELD) standards (2004, 2007, 2012) that represent the second language acquisition process. The five ELD standards cover the language students need to comprehend and produce in five areas of academic English language: social and instructional language and the language of the four content areas of language arts, mathematics, science, and social studies. Based on these standards, WIDA has developed the following assessments for use by WIDA states:

- a K–12 annual summative English language proficiency (ELP) test, Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs®);
- an initial screener, the WIDA ACCESS Placement Test (W-APT™); and

- an on-demand, “off-the-shelf” test of ELP known as WIDA MODEL™ that can be used for placement or for interim assessment both within WIDA states and also by non-WIDA educational programs.

In addition to its standards and assessments, WIDA pursues a research agenda on behalf of member states. WIDA research explores not only the validity of the assessments, but also areas of interest such as ELP growth rates, correlations between its ELP tests and state academic tests, and classroom implementation of the ELD standards. Concurrently, WIDA also provides extensive professional development opportunities and maintains a comprehensive Web site (www.wida.us). The Central Office of the WIDA Consortium serves as the management partner for the ASSETS Consortium in carrying out the activities proposed in the current Enhanced Assessment Grant.

2.2 ASSETS assessment system

The funding awarded for the ASSETS project will allow the ASSETS Consortium to develop a technology-enhanced assessment system to assess the acquisition by English language learners (ELLs) of the academic English language needed for college and career success. While it will take longer than the four-year grant period to realize the complete vision of a national language assessment system that fully leverages technology, at the end of the 2014-15 school year the ASSETS assessment system will be a comprehensive system that is (a) technology-based, incorporating several major technological enhancements; (b) anchored in the established English language development (ELD) standards developed by the WIDA Consortium, which are aligned with the Common Core and other state content standards; (c) informed by ongoing research; and (d) supported by comprehensive professional development and outreach, all of which will be developed within the framework of the multistate ASSETS Consortium.

The ASSETS assessment system will include:

- **English Language Development (ELD) standards:** WIDA’s English Language Development Standards will be a key component of the ASSETS assessment system. First published in 2004 and updated in 2007 and 2012, the standards capture an evolving understanding of (a) the language needs of ELLs and their educators and (b) the use of the standards as the foundation for instruction and assessment.
- **A common definition of English language learner:** An ASSETS Consortium subcommittee will develop a common definition of English language learner that will be operationalized in the assessment system. A common home language survey will be developed to be used as a tool in the initial identification of ELLs, and Consortium states will agree to recognize a common score on the screener and annual assessment as one indication that a student no longer qualifies for language support services.
- **English Language Proficiency (ELP) assessments:** The ELP assessments will use technology to allow for (a) more authentic language assessment tasks for all language domains (Listening, Speaking, Reading, and Writing); (b) timelier score reporting; (c) reduced burden on test administrators; and (d) compatibility to the greatest extent possible with content-driven

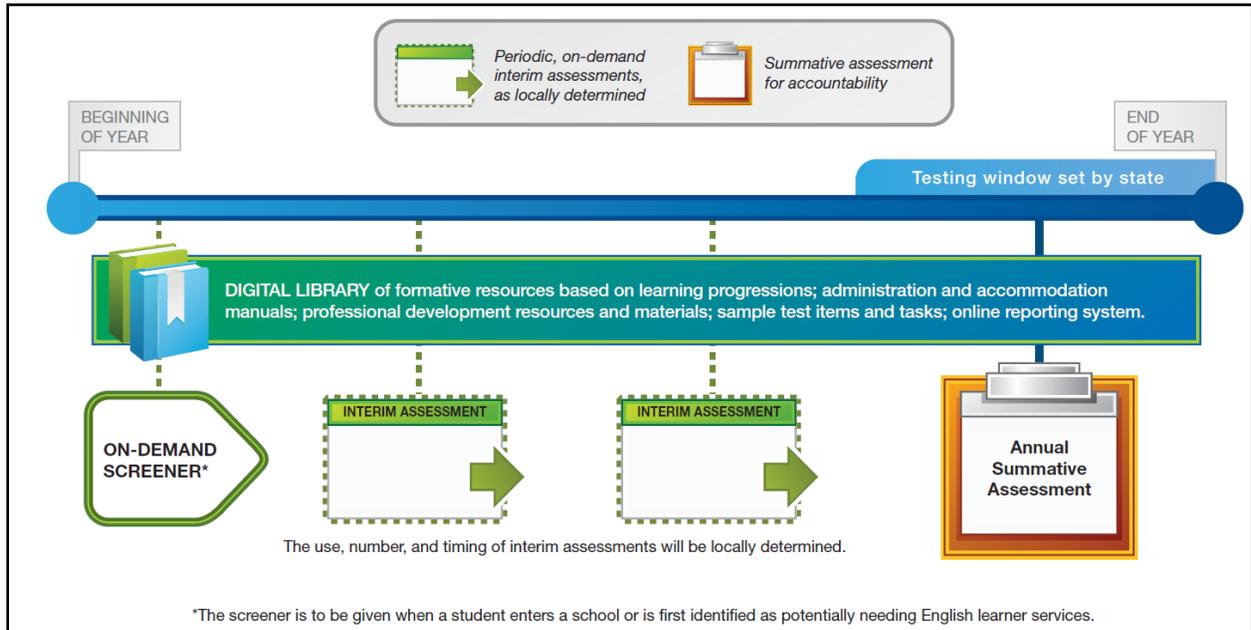
assessment systems, including those of the Partnership for the Assessment of Readiness for College and Careers (PARCC), the SMARTER Balanced Assessment Consortium (SBAC), Dynamic Learning Maps (DLM), and National Center and State Collaborative (NCSC), as well as with individual state achievement measures. The new assessments to be developed are as follows:

- **A computer-based summative test, ACCESS 2.0:** ACCESS 2.0—to be administered annually in grades K–12¹ for accountability and program purposes—will cover (a) the language domains of Listening, Speaking, Reading, and Writing and (b) the five WIDA ELD Standards, encompassing social and instructional language and the language of language arts, mathematics, science, and social studies.
- **A computer-based on-demand diagnostic (screener) test, the ACCESS 2.0 Screener:** The ACCESS 2.0 Screener may be used to determine eligibility for ELL services and program placement within those services. The test format will be derived from that of ACCESS 2.0, but all scoring not done automatically by computer will be done locally.
- **Computer-based interim assessments:** A series of shorter, targeted interim assessments will be developed to enable schools to chart student progress along the English language development continuum in finer increments and with more precision than ACCESS 2.0 allows and to help guide instruction. Innovative, technology-enhanced item types will be piloted within the interim assessments and transitioned to ACCESS 2.0 as appropriate. The combination of partial-credit scoring and analysis of patterns across responses will be used to enhance the diagnostic value of the immediate feedback enabled by the computer delivery of the assessments.
- **Formative assessment resources:** English language learning progressions will be researched and developed for both the academic and social English associated with academic success and career readiness. These progressions will provide a foundation for the development of formative assessment resources to help educators monitor student growth during instruction.
- **A training program for local scorers:** A scorer training program based on the Center for Applied Linguistics' existing Multimedia Rater Training Program (MRTP) will provide intensive, on-demand training and practice in scoring Speaking and Writing performances on the screener and the interim assessments.
- **Professional development and outreach materials:** Materials and methods will be developed for (a) professional development on how to implement the ASSETS comprehensive assessment system, including the appropriate and effective use of assessment results, and (b) outreach to stakeholders, including families, policymakers, and researchers.

¹ While the kindergarten test is part of the planned assessment system, its development is not funded under the ASSETS grant.

Figure 3 shows how the components of the ASSETS assessment system relate to one another.

Figure 3. The ASSETS assessment system

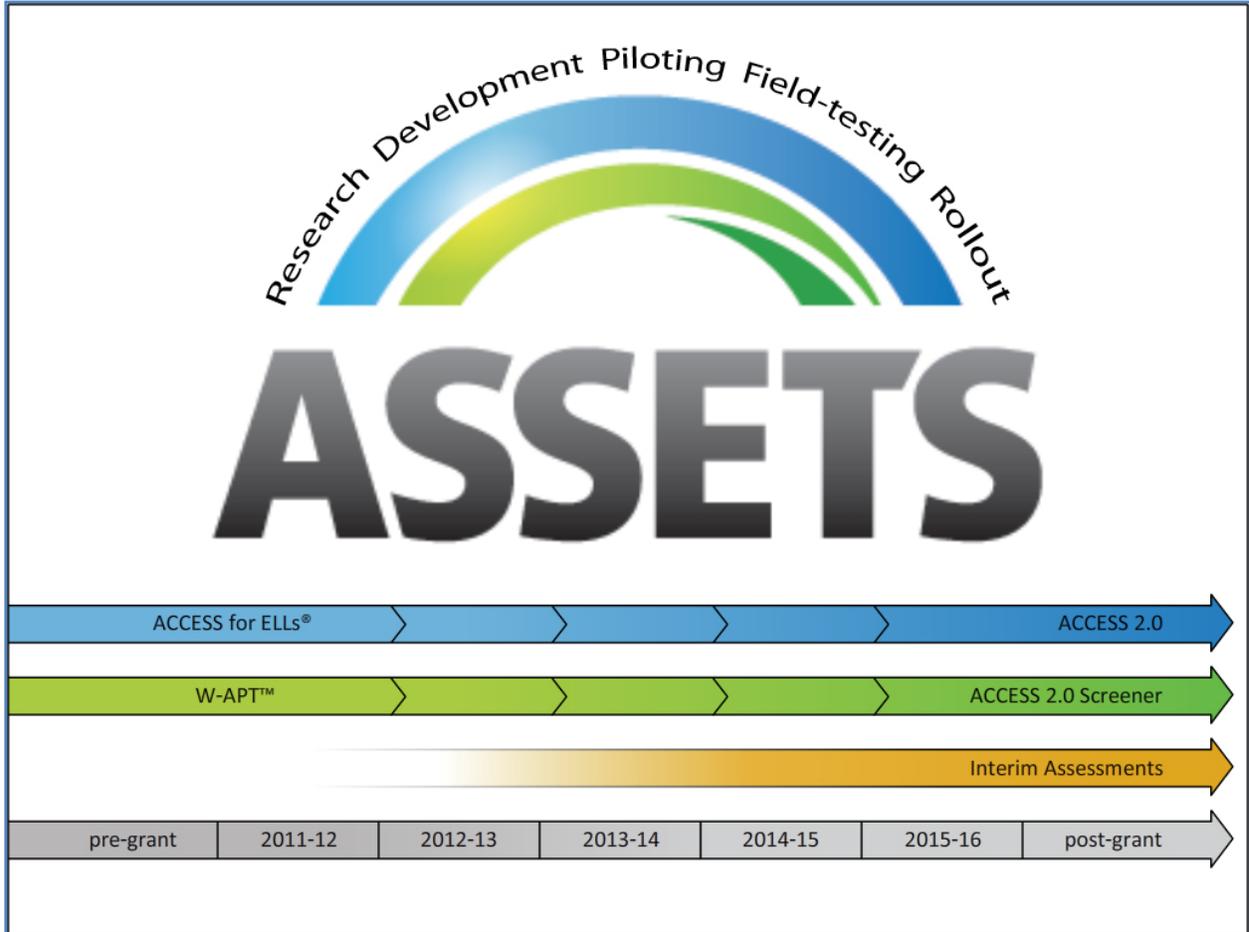


(http://www.k12center.org/rsc/pdf/18901-ASETSConsortiumFlier_WEB.pdf)

[N.B. CAL and WIDA are collaborating on development of a version of the above diagram that more accurately represents the ASSETS assessment system as currently envisioned.]

While the ASSETS assessment system is comprehensive, the funding awarded to the ASSETS Consortium does not cover the full development of all components of the assessment system. The complete development of the technology-based annual summative assessment (ACCESS 2.0) and the on-demand diagnostic screener (the ACCESS 2.0 Screener), however, is covered, as is the research and development of innovative technology-enhanced interim assessments for selected grades. Figure 4 shows the ASSETS project using the funding award to bridge the transition from the existing ACCESS for ELLs® and W-APT™ to the computerized ACCESS 2.0 and ACCESS 2.0 Screener.

Figure 4. Assessment system components funded by the ASSETS grant



2.3 Levels of documentation

This document presents more detail about the annual summative and screener assessments outlined in broad strokes in *The ASSETS Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener*. Both the *Assessment Framework* and the *Test and Item Design Plan* are public documents geared primarily for representatives of State Education Agencies (SEAs) to use as consensus-building tools among Consortium members and as documentation and communication comprehensible to a general audience involved in the assessment of ELLs.

In addition to these consensus-building documents, the Center for Applied Linguistics (CAL), the main test development partner, and the WIDA Consortium plan to produce a series of academic research papers and reports. This academic paper series will explicate the theoretical foundations underlying the assessment system (the construct of Academic English, for example), document the findings of qualitative and quantitative research conducted during the various phases of its development, and

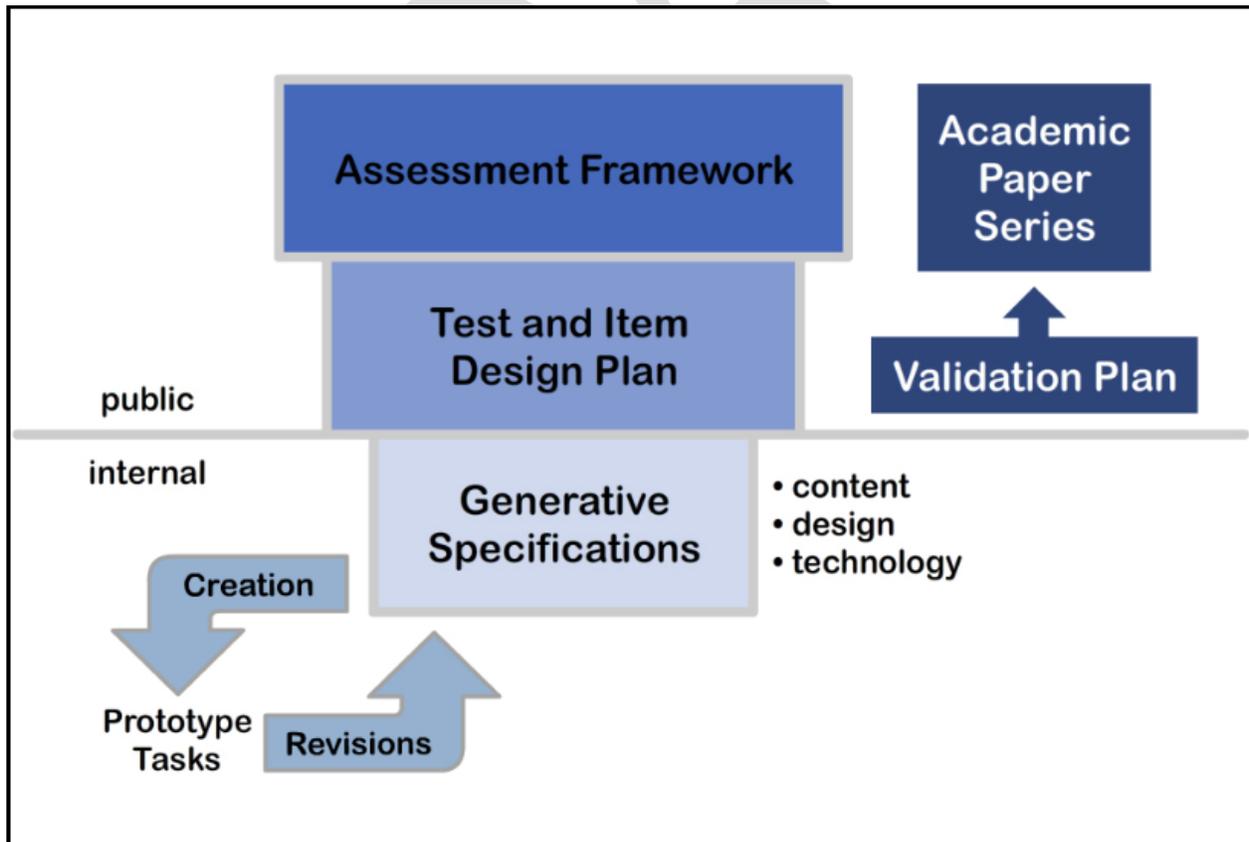
provide technical psychometric analyses that constitute support for the use of the system and its components.

The selection of topics to be covered in the research papers and reports will be guided in part by a validation plan. The validity of a test is the extent to which evidence supports the interpretations of test scores associated with the proposed uses of the test, and the ACCESS 2.0 validation plan will be a document outlining the argument from evidence that needs to be made in order to justify the intended uses of ACCESS 2.0 and the ACCESS 2.0 Screener.

Based on the agreed-upon details contained in the *Framework* and the *Test and Item Design Plan*, CAL will develop for its internal use detailed generative test and item specifications. These specifications, covering all details of test and item formats, including content, design, and technology specifications, will guide the development of the initial prototype assessment tasks to be researched and developed as part of this project. Research conducted on these prototypes will either confirm the specifications or inform modifications to them in an iterative fashion during the piloting stage of test development. The final version of these generative specifications will guide the development of all items needed during the field testing stage of test development.

These levels of project documentation are represented in Figure 5.

Figure 5. Levels of project documentation



2.4 ACCESS and ACCESS 2.0

As their names suggest, ACCESS 2.0 and the ACCESS 2.0 Screener have as their foundation WIDA's K-12 annual summative English language proficiency test, ACCESS for ELLs®. ACCESS 2.0 is not, in other words, being developed from scratch, but is, rather, building on a respected and widely-used assessment.

ACCESS for ELLs® and ACCESS 2.0 have many features in common:

- Each item or task targets at least one of the five foundational WIDA English Language Development Standards: Social and Instructional Language (SIL), Language of Language Arts (LoLA), Language of Math (LoMa), Language of Science (LoSc), and Language of Social Studies (LoSS).
- The tests are comprised of four domain-specific subtests: Listening, Reading, Writing, and Speaking.
- Test takers are evaluated according to three performance criteria, one at each of the discourse (linguistic complexity), sentence (language forms and conventions), and word/phrase (vocabulary usage) levels.
- A test taker's score is mapped onto one of five language proficiency levels: "Entering," "Emerging," "Developing," "Expanding," and "Bridging." (The "ceiling" of English language proficiency defined by the WIDA standards for assessment purposes is called "Reaching.")
- What language ELLs will process, understand, produce, or use at each of the five defined language proficiency levels is specified in performance definitions that address desired linguistic attainments at the three levels of language analysis: discourse, sentence, and word/phrase.
- Within each combination of grade-level cluster, standard, and language domain, Model Performance Indicators (MPIs) found in or developed from the WIDA ELD Standards describe the expectations for ELLs at each of language proficiency levels 1-5 and are operationalized on the test. The sequence of five MPIs together describes a logical progression and accumulation of skills on the path from the lowest level of English language proficiency to full English language proficiency for academic success.

More information on these features of the ACCESS tests new and old can be found in *The ASSETS Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener*, on the WIDA (www.wida.us) or ASSETS Web sites (www.assetsproject.org), or in the appendix at the end of this document. Table 2 lists some ways in which ACCESS 2.0 and the ACCESS 2.0 Screener differ from past incarnations of the ACCESS test and indicates where the reader can find out more.

Table 2. Major features of ACCESS 2.0 (summative and screener)

Feature		Where to find more information
The summative test and screener are part of a much larger assessment system.		TIDP 2.2
Extensive documentation will be produced.		TIDP 2.3
The tests will be aligned with WIDA’s 2012 Amplification of the ELD Standards, which correspond to state content standards including the Common Core State Standards.		Framework 5.1.3
Scores by standard will be added to score reports.		TIDP 4.7.1
Grade 1 will have its own test form, Grades 2-3 will comprise a new grade cluster, and Grades 4-5 will comprise a cluster.		Framework 5.2
The test will more systematically sample the spectrum of language functions that comprise academic English language proficiency.		TIDP 4.3
ACCESS 2.0 and the ACCESS 2.0 Screener will be computer-delivered.	The Speaking subtest will not be delivered face-to-face. Speaking prompts will be delivered by computer and student speech samples digitally recorded and, for the annual summative assessment, centrally scored.	TIDP 8
	The Listening prompts will be pre-recorded and delivered by computer.	TIDP 5.3
	The Listening and Reading subtests will be automatically scored by the computer.	TIDP 5.6, 6.6
	For scoring Writing and Speaking on the Screener, local educators will benefit from a computerized multimedia scorer training program (ACCESS 2.0 Screener Rater Training Program—Speaking and ACCESS 2.0 Screener Rater Training Program—Writing) that individualizes instruction on rating to the individual learning trajectory of the educator.	TIDP 7.7.2, 8.6.2
	Test administration may require a tech coordinator as well as a test administrator.	Framework 8.2

3 ACCESS 2.0 at a Glance

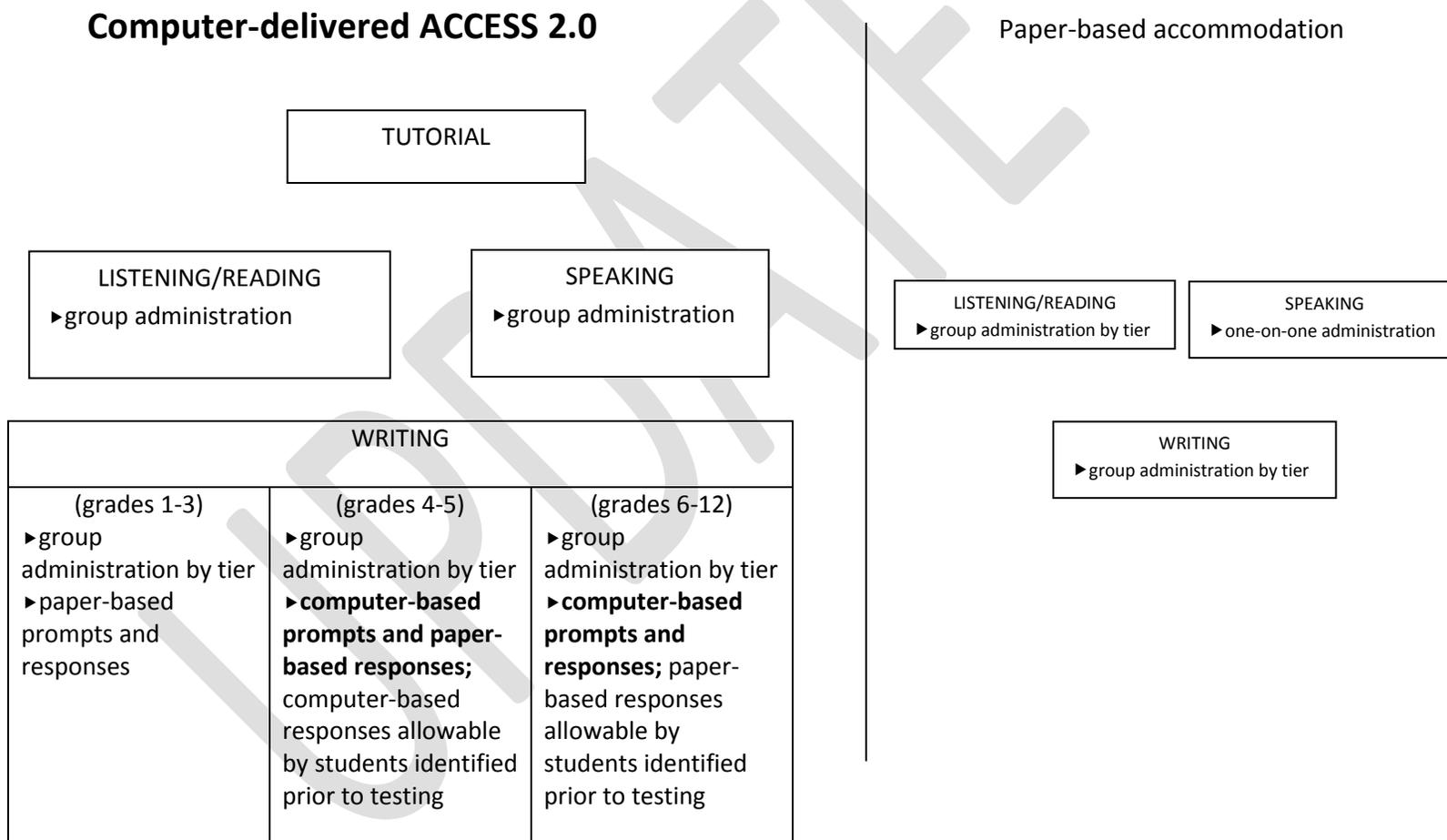
Like ACCESS for ELLs®, ACCESS 2.0 will be comprised of four subtests, one in each of the four language domains: Listening, Reading, Writing, and Speaking. There will be three separate administrations, one for Listening and Reading and one each for Writing and Speaking. For reasons explained in the section on adaptivity (4.4), the sequence of administration is: Listening and Reading in that order followed by Writing and Speaking in whatever order is logistically easier. Except for special cases—students requiring the paper-based accommodation—ACCESS 2.0 will be computer-delivered. Students in grades 1-3, however, will take the Writing subtest entirely on paper. Students in grades 4-5 will see computer-delivered prompts for the Writing subtest but will handwrite their responses unless they are identified in advance as needing to keyboard their responses. Students in grades 6-12 will see computer-delivered prompts and keyboard their responses unless they are identified in advance as needing to handwrite

their responses to writing prompts (see 7.5). All subtests of ACCESS 2.0 can be group administered and, with the exception of Writing in grades 1-3, administrations can include students in different tiers. Each subtest of ACCESS 2.0 will include an introduction and practice questions. These features of ACCESS 2.0 are summarized in Figure 6.

UPDATE

Figure 6. ACCESS 2.0 at a glance

ACCESS 2.0



4 Test-level Concerns

4.1 Design and functionality

4.1.1 General principles

ACCESS 2.0 will not simply be ACCESS for ELLs® presented on a computer screen. Like ACCESS for ELLs®, though, ACCESS 2.0 will be designed to facilitate test takers' efforts to demonstrate what they can do with the English language. Decisions regarding the look and functionality of the test, therefore, will be made with the following considerations in mind:

- **simplicity and consistency:** To minimize distraction and maximize ease of use, both the layout and the functionality of the test will be as uncomplicated and predictable as possible. The test will look uncluttered and function intuitively for the test taker.
- **student comfort and engagement:** The test will be welcoming to students and put them as much at ease as is possible within the context of a high-stakes test.
- **access to necessary content and supports:** Test takers will see or be able to easily access what they need to perform to their potential on a given item or task.
- **construct fidelity:** Transferring ACCESS to the computer will not change what is being tested. Computerizing the test will involve identifying which features of the test are artifacts of the current test's paper/in-person delivery or of the computerization process, and which are construct-relevant, in order to minimize construct-irrelevant sources of differences between student performances.
- **developmental differences:** To best serve all test takers, the look and functionality of the test will differ across grade-level clusters. Careful thought will need to be devoted to the specifics of making each computer-based test age-level appropriate.
- **accessible, non-biased items:** Acknowledge the diversity of test-takers, particularly ELLs with disabilities, by developing items using the principles of Universal Design and computer-embedded supports to meet the specific access needs of students. Carefully review items for possible bias and cultural insensitivity with this diverse group of test-takers.

The general considerations above translate into more concrete principles, such as the following:

- **Use a small, fixed number of layout templates.** Navigation components will always appear in the same place on the screen. Stimulus pictures and text, item stems, and response options will appear in predictable locations, with limited variation allowed to accommodate differences in text length, number of response options, and degree of graphic support.
- **Preserve the thematic folder concept familiar from ACCESS for ELLs®.** On the current ACCESS test, items (Reading and Listening) and tasks (Writing and Speaking) are presented in thematically linked sets called "thematic folders" or just "folders." Presenting items and tasks in this way reduces the number of unique contexts to which test takers must orient themselves.
- **Include on the screen with the question all the information students need to answer it.** Students should not have to toggle between screens as they work through an item or task. If it is

necessary to look at a graph to answer a question, then the graph, the question, and the response space will appear together on one screen. If questions pertain to a reading passage, then both the passage and questions will appear on a split-screen.

- **Provide students with an indication of their progress through the test.** As students work through each domain subtest, they will want to know how much of the test remains to be completed. Test takers are accustomed to being able to gauge their progress based on the number of pages left in the test booklet.
- **Provide multiple presentation and response formats for test questions,** including a paper test alternative for the small number of students who are unable to use a computer.
- **To the greatest degree possible, display information in a flexible format** so that perceptual features such as the following can be varied as per student need:
 - The size of text, images, graphs, tables, or other visual content
 - Color used for information or emphasis
 - The volume of speech or sound
 - The speed or timing of video, animation, sound, simulations, etc. (if used).
- **Standardize the interoperability and accessibility of computer-based test items.** The test will incorporate the Accessible Portable Item Profile (APIP) interoperability framework to allow items to be tagged so they look and function the same regardless of the browser used; and so that information about a student's access needs and accommodations are embedded in individual test items and within the test itself (e.g., zoom or speech-to-text).

Domain-specific interface and layout issues will be addressed in later sections of this document.

4.1.2 Design elements

To achieve consistency across domains and general ease of use, certain elements will appear throughout the ACCESS 2.0 test. Other elements, though they may change form from one domain to another, may occupy a standard position on the screen. Table 3 shows a *possible* set of test-wide design elements. Their numbers may be reduced or augmented as additional decisions are made about how the test should look and work.

Table 3. Test-wide design elements

ELEMENT		DESCRIPTION
navigation bar		A test navigation bar will appear at the bottom of the screen. This bar will remain fixed throughout the test, and navigation buttons will appear or not as needed.
navigation bar	go button	 The go button will appear on the navigation bar on the introductory screen of a thematic folder. Students will click the go button to begin the first item or task in the folder. The go button serves as an additional indicator—above and beyond the theme graphic—that a new item/task is beginning.
	forward button	 The forward button will allow students to submit their answers and proceed to the next screen.
	progress bar	The progress bar will indicate progress through a domain subtest. (Note that the status of this element is uncertain.)
	question number	 Question 1 The question number will indicate the question number within a domain subtest. (Note that the status of this element is uncertain.)
item input space		Part of the screen layout will be designated space for item input (e.g. graphics, reading texts). The look and content of the item input space will be defined by domain.
student workspace		Part of the screen layout will be designated space for student responses (e.g., selected response options, space for writing responses). The look and content of the student workspace will be defined by domain.

4.1.3 Navigation

Students accustomed to being able to revisit test questions on a paper-and-pencil test may experience anxiety if they are unable to do so on a computer-delivered test. There are compelling reasons, however, for not allowing backward navigation on ACCESS 2.0. While the proposed folder-level adaptivity (see section 4.4) precludes students revisiting a folder once it has been completed, examinees could, in theory, be permitted to revisit items *within* a folder. But communicating to examinees that they may revisit items only within a folder may confuse test takers, particularly young ones or those with low English proficiency. And allowing students to go back to previous questions may also encourage them to think that they *should* go back, that information on earlier screens could help them answer the question at hand. (Each ACCESS 2.0 item, though thematically related to the others in its folder, is independent of them.) If examinees are not allowed to return to a question once it is answered,

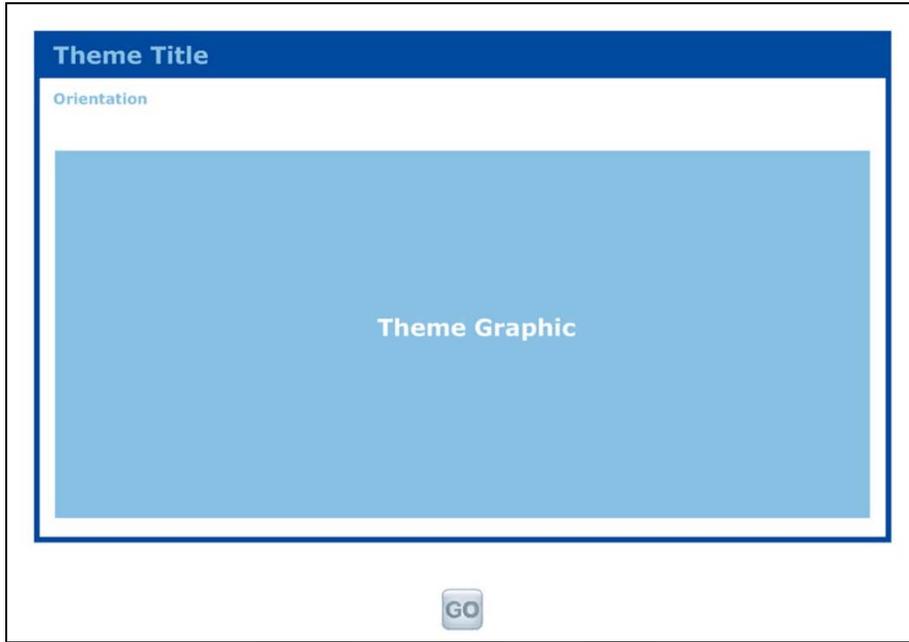
however, they will neither waste time navigating backward and forward nor agonize over whether they should.

For the reasons outlined above, a simple, one-button navigation scheme will be implemented. There will be a go button on the introductory screen of each thematic folder. Students will press it to begin answering the questions within the folder. On each subsequent screen within a folder, the go button will be replaced by a forward button. If a question appears on a screen, pressing the forward button will both submit the answer entered and take the test taker to the next screen. To guard against students unintentionally skipping questions, test takers trying to leave a screen without answering the question on it will be unable to do so, and will need to select a response before moving on. Figure 7 (a and b) shows the location of the navigation buttons.

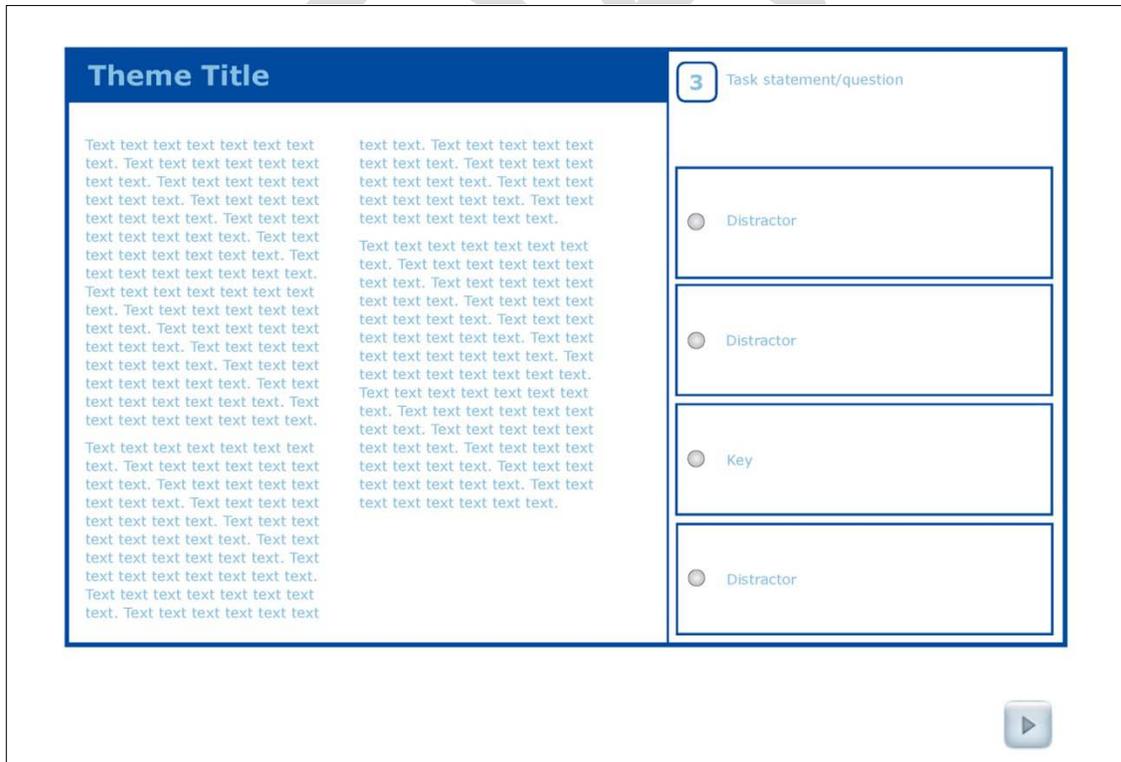
UPDATE

Figure 7.

a.



b.



4.2 Accessibility and affect

4.2.1 Instructions

Students taking ACCESS 2.0 should never be confused about what they are expected to do or uncertain about how to interact with the test. Every effort will be made, therefore, to ensure that test takers at all levels of English language proficiency understand the test instructions, all of which will be in English. Instructions will be written, for one thing, to be comprehensible to students at the lowest proficiency levels. Also, instructions will be

seen—shown on the screen—and heard—“read aloud” when examinees arrive at the screen in question. Graphical support will also be provided as appropriate to the targeted proficiency level of the test question.

Note, however, that, as shown in Table 4, when Reading is being tested, test stimuli and questions will be text only. Likewise, when Listening is being tested, test stimuli and questions will be audio only. Comprehension of written text is part of the construct of Reading, while comprehension of spoken text is part of the Listening construct. Nevertheless, graphical support may be supplied as appropriate to the targeted level of proficiency.

For Writing, the specific task prompt (i.e., the prompt to which the student’s writing is a response) will be shown as text and read aloud. For Speaking, all parts of the task will be presented in both text and audio formats. This includes instructions, the task input, and the task prompt (i.e., the input to which the student’s response is a rejoinder).

Table 4. Means of delivery of item/task components

Domain	Component	Means of Delivery
READING	instructions	text and audio, with graphical support (illustrations and photographs) as appropriate
	stimulus	text, with graphical support as appropriate
	question	text
	response options	text and/or graphics
LISTENING	instructions	text and audio, with graphical support as appropriate
	stimulus	audio, with graphical support as appropriate
	question	audio
	response options	graphics and/or text
WRITING	instructions	text and audio, with graphical support as appropriate
	stimulus	text and audio, with graphical support as appropriate
	task prompt	text and audio
	response	student-generated

SPEAKING	instructions	text and audio, with graphical support as appropriate
	stimulus	text and audio, with graphical support as appropriate
	model response	audio only
	task prompt	text and audio
	response	student-generated

4.2.2 Virtual test administrator

Within the context of the current ACCESS for ELLs® administration, adult guidance comes from three sources. There is (1) the test administrator who presides over the group administration of the Listening, Reading, and Writing subtests, and who is especially important as a support to students taking the Tier A test; (2) the administrator of the one-on-one Speaking subtest; and (3) the “narrator” who delivers instructions on the media-delivered Listening subtest (2013-14). On ACCESS 2.0 the types of adult facilitators will be reduced to two. Students will still take the test in a room supervised by a test administrator, and a cadre of *virtual* test administrators will guide the test takers’ work on the computer. While the virtual test administrator (virtual TA) may differ from one subtest to another and may be heard but not seen in domains other than Speaking, it is expected that the virtual TA will be a comforting presence across the ACCESS 2.0 practice exercises and subtests.

4.2.3 Variation across grade levels and language proficiency levels

The format of the existing ACCESS for ELLs® varies across grade-level clusters and proficiency levels. Both the look of the test and the amount of support provided changes to reflect the differing needs of test takers, both as they grow in their academic English language proficiency and as they advance in age. This same sort of purposeful variation will be carried over onto ACCESS 2.0, with the computer-delivery of the test perhaps allowing for even more tailoring. The following characteristics of the test may vary across grade-level clusters and/or language proficiency levels:

- **font/font size:** Text presented to younger students will likely appear in a larger font, and the font itself may even be different (i.e., more “child friendly”) from that used for students in the higher grades.
- **graphics:** All graphic supports—whether illustrations or photographs—will be designed to be age-appropriate, in that they will (1) show children of similar age to the test takers in situations likely familiar to test takers or (2) try to replicate to the extent possible materials students in a particular age group are likely to encounter in the classroom. This means not only that images for younger test takers will be less detailed and more “kid-friendly” than those used on the test forms for older students, but also that graphs and maps will be presented differently to students of different ages.
- **support:** Students at lower grades or language proficiency levels will likely be given more support than those who are older or more proficient. This additional support might take the form of modeling sample responses or extensive use of graphics.

- **student control:** Cognitive labs during the pilot phase will explore the possibility of allowing older students more control over their progress through the test than that afforded to younger students. For example, older students may have more control over the pacing of the test—e.g. through decreased use of auto-progress—than younger students.

4.3 Key academic language functions

As stated in *The ASSETS Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener*, one goal is for ACCESS 2.0 and the ACCESS 2.0 Screener to sample from the breadth of social and academic English language critical for success in today’s classrooms. Test developers will ensure that all aspects of academic English language are covered by (1) identifying academic language functions students should be able to either produce or comprehend, and (2) determining how to best distribute these language functions across the test forms (i.e., grade-level clusters, standards, and domains).

The notion of language function serves as a useful heuristic to describe how language is used to engage meaningfully in academic content areas at all levels of language: discourse, sentence, and word.

The term *language function* is used to describe the language that a language user needs in order to fulfill a communicative purpose within a given context. A function refers to what a language user does with and through language. This places the focus on using language for meaning making and what students are doing with language to accomplish various communicative activities.

There are many lists of language functions, some enumerating as few as six functions and others identifying as many as 75. Some example academic language functions are *describe*, *sequence*, *generalize*, and *enumerate*. CAL and WIDA staff, in collaboration with researchers at UCLA, have identified four “overarching language functions”—Explain, Argue, Recount, and Discuss—connected to the text types discussed in the Common Core writing standards, as well as an assortment of underlying microfunctions. These language functions will be incorporated into the generative item specifications from which ACCESS 2.0 items and tasks are produced.

4.4 Adaptivity

The existing ACCESS for ELLs® consists of three overlapping tiers: Tier A (targeted to students at the lowest levels of English language proficiency), Tier B (targeted to students at the mid levels of English language proficiency), and Tier C (targeted to students at the highest levels of English language proficiency). As the WIDA Web site explains, “this keeps the test shorter and more appropriately targets each student’s range of language skills.” What the current tier system allows ACCESS to do, folder-level adaptivity will allow ACCESS 2.0 to do even better. Note that, while the ACCESS 2.0 Screener will comprise fewer folders than ACCESS 2.0 itself, adaptivity will work the same on both tests.

The proposed plan for adaptivity (see sections 5.2, 6.2, 7.2, and 8.2) was conceived with three primary criteria in mind:

- **student affect**
 - All test takers, regardless of English language proficiency level, should enjoy the boost in confidence that comes from encountering test items they can do.
 - Exposure to items targeting proficiencies much higher than a student’s current level should be minimized, to avoid discouragement.
 - The test should not be longer than it needs to be, lest it fatigue students.
- **psychometric properties of the test**
 - The test must include enough items to allow for legitimate reporting of all promised subscores.
 - Scores on ACCESS 2.0 should be at least as psychometrically precise as ACCESS for ELLs®.
- **coverage of the WIDA ELD standards**
 - The test must include enough items per standard to allow for the calculation of subscores by standard.
 - Roughly the same number of folders should be devoted to each standard on ACCESS 2.0 as on ACCESS for ELLs®.

Information about domain-specific adaptivity mechanisms will be provided in later sections (5.2, 6.2, 7.2, 8.2) of this document, but note here that simply moving ACCESS to the computer yields some tier-related benefits. Examinees taking the computer-based test will no longer be locked into a single tier across all domains, for instance, and, at least for the majority of students (i.e., those not taking the paper-based accommodation), educators will no longer need to preorder tiered tests.

4.4.1 Order of subtest administration

One feature of the proposed adaptivity scheme is for tier placement on the Writing and Speaking subtests to be informed by a student’s performance on the Listening and Reading subtests (which will be scored in real time). For this to be possible, of course, a student must have completed both the Listening and Reading subtests before sitting down for either the Writing or Speaking one. The sequence of administration is: Listening and Reading in that order followed by Writing and Speaking in whatever order is logistically easier. Note that while the Listening and Reading tests need not be taken as a unit, the recommended default is to administer these subtests one after the other.

4.5 Tutorial

While every effort will be made to make ACCESS 2.0 as user-friendly and self-explanatory as possible, students may need to be introduced to key features of the computerized test before it is administered to them. The practice exercises that will precede each subtest of ACCESS 2.0 will both expose students to the types of items or tasks they will encounter and give test takers a chance to familiarize themselves with any domain-specific functionality (recording controls for Speaking, for instance). To provide the opportunity for students to gain experience with the test’s functionality prior to taking the test operationally, however, a tutorial will be developed to orient students to the computer interface and give them the chance to try out their understanding of the functionality under controlled circumstances. If students know and are comfortable with how the test works, they will be able to demonstrate their current level of academic English language proficiency.

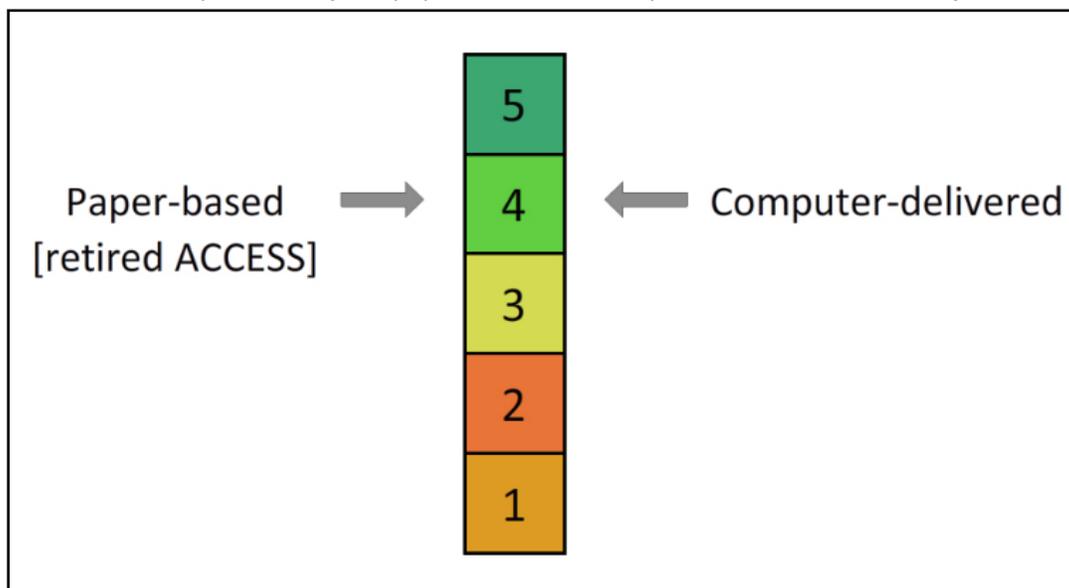
The ACCESS 2.0 Tutorial

- **will introduce students to the interface, general test functionality, and the item and task types for each domain.** The tutorial will include, for example, questions that require students to record their response (used for all students in the Speaking domain).
- **may be administered to the same student more than once.** The tutorial will be available for administration throughout the school year to familiarize test-takers with the interface prior to administration of ACCESS 2.0.
- **can serve as a tech check for schools.** Just as the tutorial will allow students to acquaint themselves with ACCESS 2.0 under low-stakes conditions, so too will it provide schools an opportunity to verify in advance of the ACCESS 2.0 testing window that they have met the test's technology requirements.
- **can help schools work out the logistics of administering a computer-delivered test.** Having the tutorial as a dry run may help ensure that everything runs smoothly at the time of actual test administration.

4.6 Paper-based accommodation

A paper-based accommodation will be available—for both ACCESS 2.0 and the ACCESS 2.0 Screener—for those few individual students who may not be able to take a computer-based test. Rather than being newly developed, however, the paper-based test will be comprised of items from retired forms of the ACCESS for ELLs® test. The paper-based and computerized forms will not be formally parallel. For example, the Speaking section of the paper-based accommodation will be administered face-to-face and scored during administration with a holistic rubric, while on the computer-delivered test students will be given their prompts and record their answers on the computer, for later scoring with a more detailed rubric. Nevertheless, psychometric equating will ensure that performances on both forms will be equivalent in terms of the interpretation of the performances on the five levels described in the WIDA standards. Figure 8 illustrates this equivalence.

Figure 8. Common interpretation of the paper-based and computer-delivered versions of ACCESS 2.0



It is important to note here that there will in fact be three ways to take ACCESS 2.0:

1. **Entirely on the computer:** In this scenario, test takers see item and task prompts onscreen and input their answers directly into the computer, whether by selecting a response option for Listening and Reading, recording their spoken responses, or keyboarding their written responses.
2. **On the computer except for handwritten Writing responses:** Pre-identified students, whose keyboarding skills are not fully developed, will take the computerized test but write their responses to Writing prompts—still presented onscreen—on paper rather than keyboarding them. Students who handwrite their responses will do so in a pre-ordered, pre-labeled test booklet. Since the blank response booklet will be generic, pre-ordering will not have to be done by tier. (Note that all students in grades 1-3 will take a paper-based Writing subtest.)
3. **Entirely on paper:** There will be a paper-based test form available for students who need it as an accommodation.

[N.B. More information about other accommodations will be added to this document in reflection of the work of the ASSETS accommodations subcommittee.]

4.7 Reporting

4.7.1 Scores reported

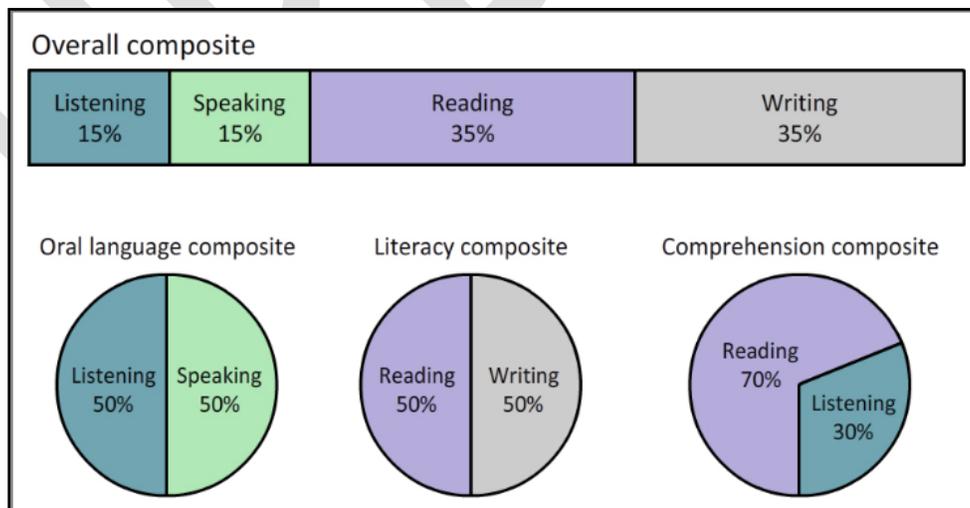
4.7.1.1 ACCESS 2.0

One goal of the ASSETS project is to include in the ACCESS 2.0 score reports scores not only by domain, as is done now for ACCESS for ELLs®, but by standard as well. Psychometric work, in combination with careful descriptions of the domains and standards, will ensure that these scores are valid and reliable. Stakeholders will be consulted to determine what interpretive information would be helpful to be able to derive from these additional subscores.

Student data from ACCESS 2.0 will therefore include but not necessarily be limited to:

- **scale scores** on a K-12 vertically aligned scale
 - **by domain** (Listening, Reading, Writing, and Speaking) and
 - **by standard** (SIL, LoLA, LoSS, LoMa, and LoSc);
- grade-level specific **interpretive proficiencies**
 - **by domain**, which characterize a student’s performance in terms of the proficiency levels 1 through 5 defined in the WIDA standards; and
- **composite scores** as diagrammed in Figure 9.

Figure 9. Composite scores



[N.B. While the timeframe for the delivery of official score reports including all of the abovementioned scores will depend on the pace of the centralized scoring of Writing and Speaking tasks, the reporting

subcommittee will explore the possibility of releasing unofficial results from the computer-scored Reading and Listening subtests earlier for use in local decision-making.]

4.7.1.2 The ACCESS 2.0 Screener

The ACCESS 2.0 Screener will be shorter than ACCESS 2.0 (for details, see sections 5.7, 6.7, 7.7, and 8.6), and the overall English language proficiency level score will be the most psychometrically valid piece of data derivable from a student’s performance on the test. Initial proficiency-level scores for each domain may prove helpful in determining students’ extent of English language support. The Screener will not produce individual standards scores.

4.7.2 Score reports

4.7.2.1 ACCESS 2.0

Score reports generated for ACCESS 2.0 will build on WIDA’s seven-year experience with delivering meaningful, uniform score reports customized to the needs of various stakeholders. Although the specific score reports to be produced will be determined by the ASSETS Consortium member states, it is expected that score reports will target audiences similar to those receiving ACCESS for ELLs® score reports. Currently, WIDA provides for the ACCESS assessment the reports listed in Table 5.

Table 5. Score reports

Report	Description
parent/guardian	<ul style="list-style-type: none"> • test results presented visually and numerically • includes domain, oral language, literacy, comprehension, composite English language proficiency level and scale scores
teacher	<ul style="list-style-type: none"> • more detailed than parent/guardian report • includes scale scores for all domains and combinations of domains and raw scores for each of the WIDA ELD standards
student roster	<ul style="list-style-type: none"> • for teachers and administrators • gives overview of English language proficiency levels and scale scores for all domains and composite scores for ELLs in a school
school frequency	<ul style="list-style-type: none"> • for teachers and administrators • shows distribution of ELLs according to their language proficiency levels for each domain and combination of domains in a school
district frequency	<ul style="list-style-type: none"> • provides the same information as the school frequency report, just for an entire district

In collaboration with SEAs and LEAs, WIDA is already providing translations of the parent/guardian report in more than 30 languages and will continue this practice with the ASSETS assessments.

While the target audiences of the above reports are not anticipated to change, the format of the reports may be altered to more clearly communicate to individual groups of stakeholders the information they most need and want. While the reporting subcommittee will make recommendations on whether the

100-600 scale used for ACCESS for ELLs® will be retained or whether a new scale will be instituted to differentiate ACCESS scores from ACCESS 2.0 scores, different ways of presenting such information as the confidence bands associated with scale scores and the relationship between scale scores and proficiencies will be explored. Input from focus groups, psychometricians, and design professionals will help ensure that score reports are maximally comprehensible and useful for their intended audiences. *[N.B. The ASSETS Consortium is creating a reporting subcommittee that will help inform decisions made about the reporting of scores and how best to ensure their interpretability and usefulness.]*

4.7.2.2 The ACCESS 2.0 Screener

For the ACCESS 2.0 Screener, a technology-based administrator interface will be developed to streamline and standardize the process of reporting results. The test administrator will enter a student's Speaking and Writing scores into the computer, where they will be combined with the Listening and Reading scores already stored there. The computer will do the number-crunching formerly left to the administrator and then generate a printable score report. Besides easing the burden on the test administrator, digitizing results in this way will allow for aggregation of student data and monitoring of screener use.

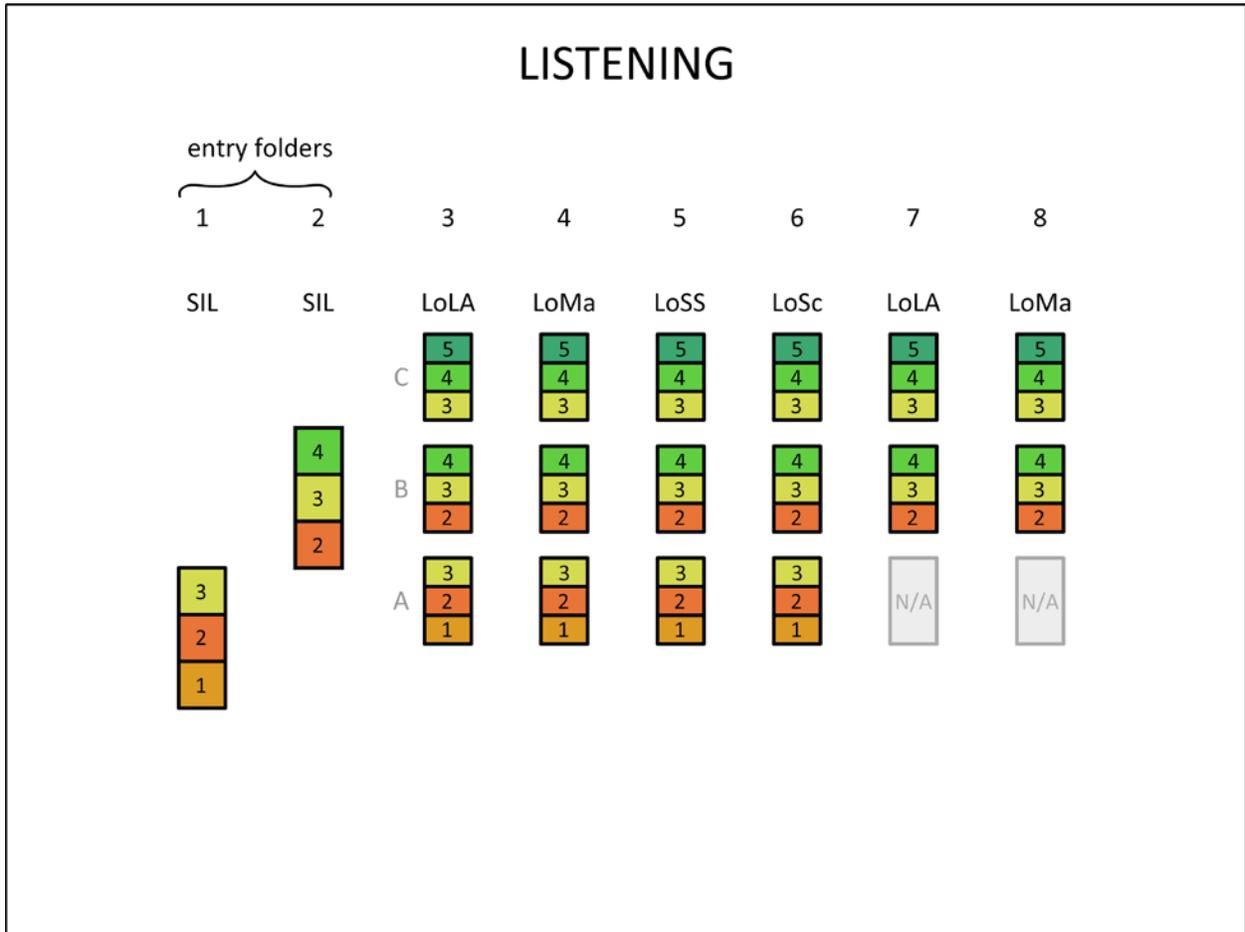
5 Subtest Specifics: Listening

5.1 Composition of the Listening subtest

Figure 10 shows the proposed structure of the Listening subtest. Each small rectangle represents an item, with the number in the rectangle indicating the level of the Model Performance Indicator (MPI) found in or derived from the WIDA ELD standards targeted by the item in question. Each stack of rectangles represents a thematic folder, and each numbered column headed by a standard abbreviation—SIL, LoLA, LoMa, LoSS, or LoSc—is called a *panel*. Though all but the first two panels include more than one folder, a test taker will be administered only one folder per panel, as will be explained in section 5.2. Readers familiar with ACCESS will note that the structure of the item pool mirrors that of the three tiered forms within a grade-level cluster on the current test. This relationship is highlighted by marking rows as “A,” “B,” and “C.”

All students, then, will take two folders targeting Social and Instructional Language and one folder targeting each of the academic standards. Students at the higher language proficiency levels will see *two* folders targeting each of LoLA and LoMa. Depending on a test taker's language proficiency, the Listening subtest of ACCESS 2.0 will consist of between 18 and 24 selected response items, grouped into between six and eight thematic folders.

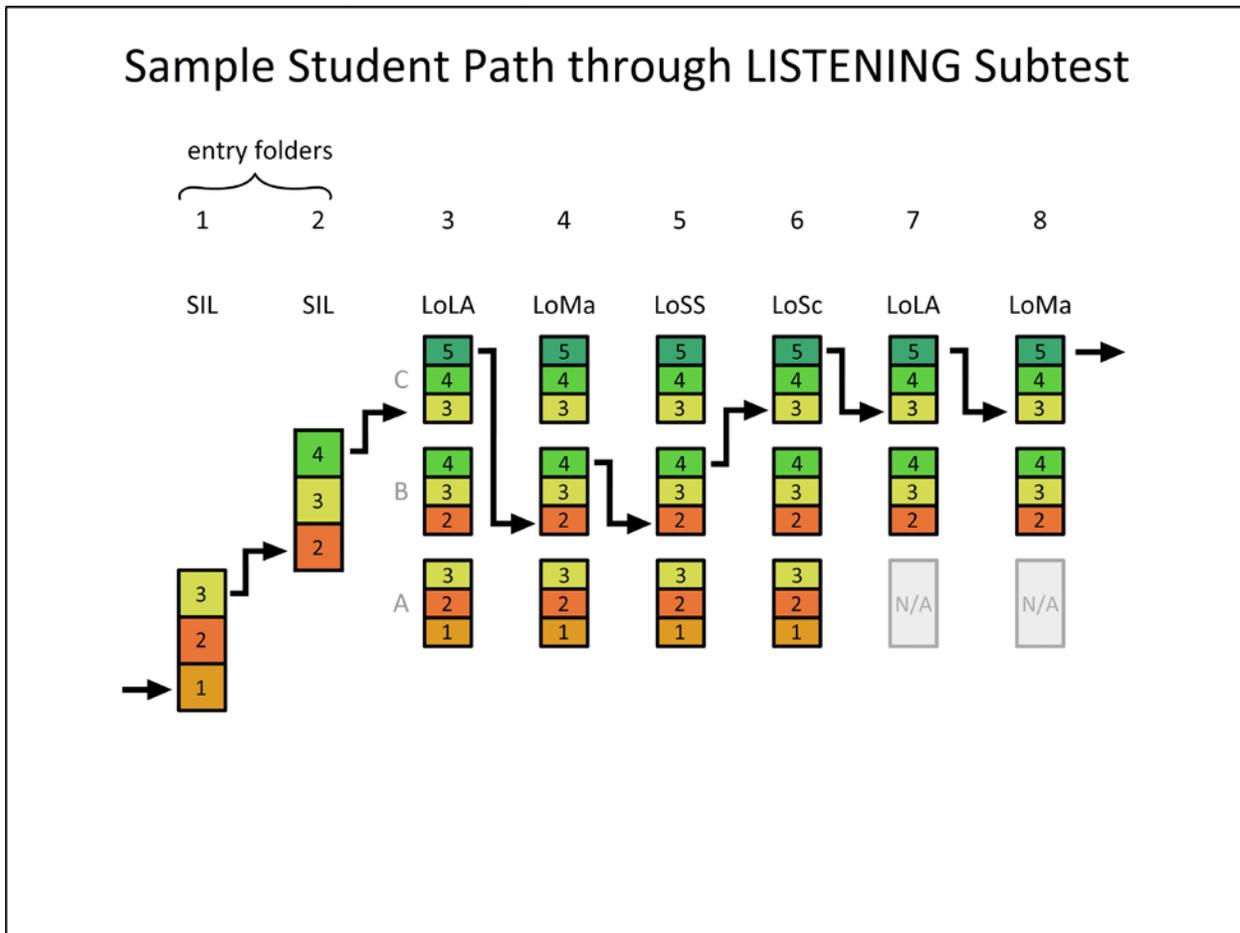
Figure 10. The structure of the Listening subtest



5.2 Adaptivity

All test takers will begin the Listening subtest of ACCESS 2.0 with two entry folders, both targeting Social and Instructional Language (SIL). The test taker's performance on the six items in the entry folders will determine which of the three leveled LoLA folders in Panel 3 s/he is administered. Throughout the test, an examinee's underlying ability measure will be recalculated after each folder s/he completes, with the tier of the next folder to be administered chosen accordingly. Students will not all see the same folders, but the order of the panels will be invariant across test takers. All students, in other words, will have the standards targeted in the left-to-right order shown in Figure 10. Figure 11 shows one path a student might take through the Listening subtest. Note that the test taker always begins a folder with the least challenging question and works up to (and exits the folder after) the most challenging.

Figure 11. Sample path through the Listening subtest



Note that it is expected that test takers who are administered primarily Tier A folders will have their English language proficiency sufficiently measured after the completion of relatively few folders. The design depicted in Figures 11 and 12 is in keeping with ACCESS for ELLs®, in which students taking the Tier A form see two fewer non-SIL folders than those administered Tier B or Tier C.

5.3 Item presentation

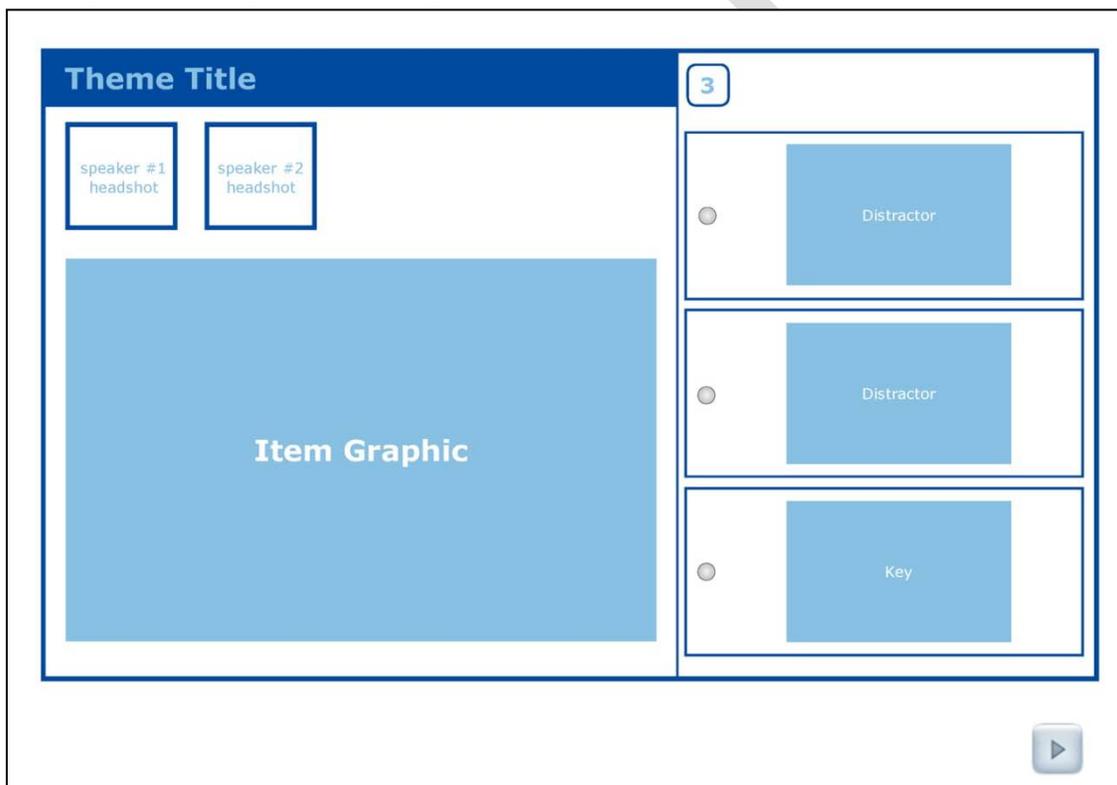
Listening stimuli on ACCESS 2.0 will be pre-recorded in a studio using voices appropriate to the item (for example, in terms of gender, age, and roles) and played back to the test taker via computer. Recorded stimuli will not only ensure a more standardized test-taking experience, but will also enable the inclusion of a broader range of interaction scenarios, including student-centered ones. A Listening stimulus, for example, might consist of an exchange between a teacher and a student as the latter asks questions to clarify a concept.

The auditory stimuli on the Listening test will be supported visually. The introductory screen of each folder will include a graphic that establishes the context of the speech students will hear in the items that follow. The graphic will show who will be speaking and the situation in which the speech will occur.

5.4 Domain-specific design elements

As indicated in the preceding section, on the Listening subtest of ACCESS 2.0 students will hear not only monologues but also dialogues. To help students keep track of who is talking as they listen to interactions among more than one speaker, each question screen will display headshots of all speakers involved in that item. Test takers will have seen the speakers in the introductory graphic used to set the context for the whole folder and will thus be able to recognize them from headshots on subsequent screens. When a character is speaking, his or her headshot will be foregrounded, either by highlighting it or by fading out the other headshot. Figure 12 shows how representation of the speakers might be integrated into the layout of a Listening screen.

Figure 12. A possible look for the Listening subtest



5.5 Student response

Students will respond to all items on the Listening subtest by selecting an option from among three choices. *[N.B. As stated in the grant proposal, alternatives to multiple choice—more technology-enhanced and performance-based item types—will be researched and developed for the lower stakes, optional interim assessments. They will be introduced onto the Reading and Listening subtests of ACCESS 2.0 and the ACCESS 2.0 Screener as appropriate, but not until the post-grant period by which time their usefulness and stability will have been researched and demonstrated on the interim assessments.]*

5.6 Scoring

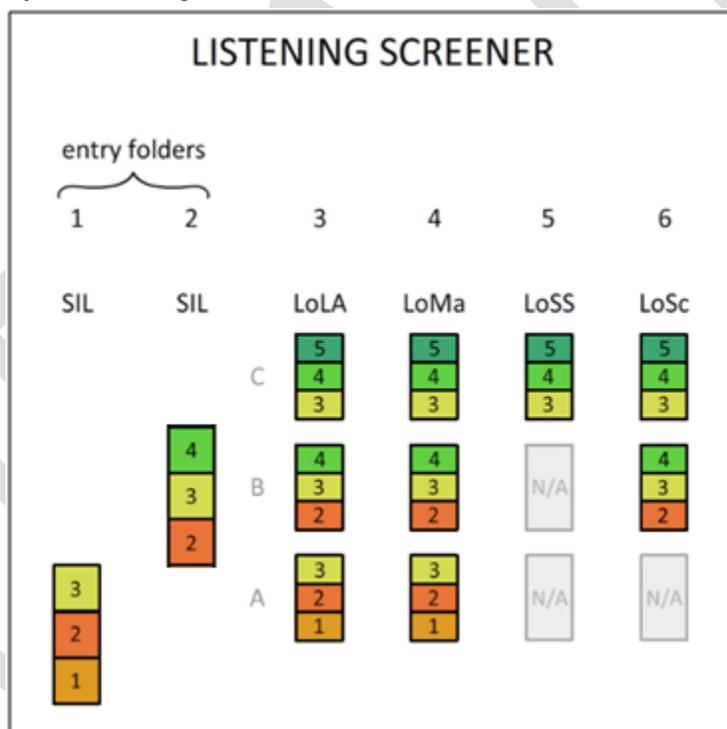
Student responses on the Listening subtest will be automatically and immediately scored by the computer. *[N.B. Score reports may not be available for the summative assessment, however, until the centralized scoring of the Writing and Speaking tasks has been completed.]*

5.7 Listening screener

The Listening screener is envisioned as a shortened version of the ACCESS 2.0 Listening subtest. Adaptivity will work on the screener as it does on the summative subtest. Students who demonstrate high levels of English language proficiency on the Listening screener will take a total of six folders, including one for each standard. Students who track at the mid or low proficiency levels, however, will finish the screener after seeing a reduced number of folders.

Note that while students taking primarily Tier B folders will not be exposed to all standards on the Listening screener, the standard they miss—Language of Social Studies—will be covered on the Reading screener (see Section 6.7). Figure 13 shows the structure of the Listening screener.

Figure 13. The structure of the Listening screener



6 Subtest Specifics: Reading

6.1 Composition of the Reading subtest

Figure 14 shows the proposed structure of the Reading subtest. In keeping with the design of the current ACCESS, the Reading subtest of ACCESS 2.0 will include more folders than the Listening subtest.

This reflects the greater weighting of Reading in the calculation of a test taker’s overall composite score: 35% Reading versus 15% Listening (see section 4.7.1.1).

All students will take two folders targeting each of SIL, LoMa, and LoLA, and one folder targeting each of LoSS and LoSc. Students at the higher language proficiency levels will see an additional folder targeting each of LoSS and LoSc (i.e., two folders for each standard). Depending on a test taker’s language proficiency, the Reading subtest of ACCESS 2.0 will consist of between 24 and 30 selected response items, grouped into between eight and 10 thematic folders.

Figure 14. The structure of the Reading subtest



6.2 Adaptivity

Adaptivity on the Reading subtest will function as described for the Listening subtest in section 5.2.

6.3 Item presentation

On the Reading subtest, texts will appear on the computer screen. Care will be taken to alleviate, to the greatest extent possible, the strain of reading on the computer screen. For each thematic folder, the folder introduction screen will provide visual support to the type of text-based material (e.g., book, newspaper, Web site) from which the text is drawn. Cognitive labs and pilot testing will inform decisions

alternatives to multiple choice—more technology-enhanced and performance-based item types—will be researched and developed for the lower stakes, optional interim assessments. They will be introduced onto the Reading and Listening subtests of ACCESS 2.0 and the ACCESS 2.0 Screener as appropriate, but not until the post-grant period by which time their usefulness and stability will have been demonstrated on the interim assessments.]

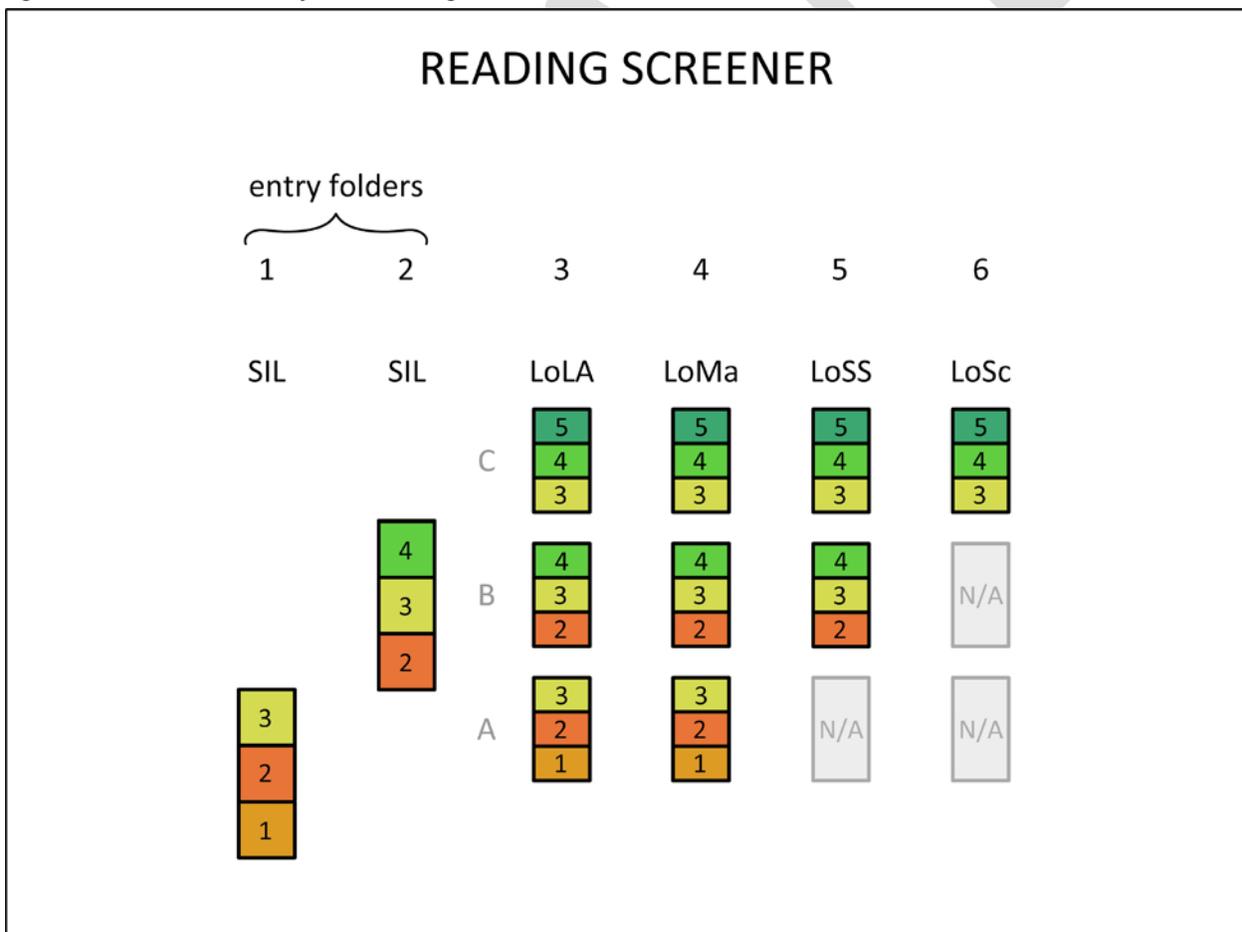
6.6 Scoring

Student responses on the Reading subtest will be automatically and immediately scored by the computer. [N.B. Score reports may not be available for the summative assessment, however, until the centralized scoring of the Writing and Speaking tasks has been completed.]

6.7 Reading screener

The Reading screener is structured similarly to the Listening screener (see section 5.7). Note that students tracking at the mid levels of language proficiency will see a Language of Social Studies folder on the Reading screener, to compensate for the lack of such a folder on the Listening screener.

Figure 16. The structure of the Reading screener



7 Subtest Specifics: Writing

7.1 Composition of the Writing subtest

The Writing subtest will comprise three tasks. On the Tier B and Tier C forms, the third task will be an extended task, as shown in Figure 17. How the standards targeted by these tasks will vary across grade levels will be spelled out in the generative item specifications.

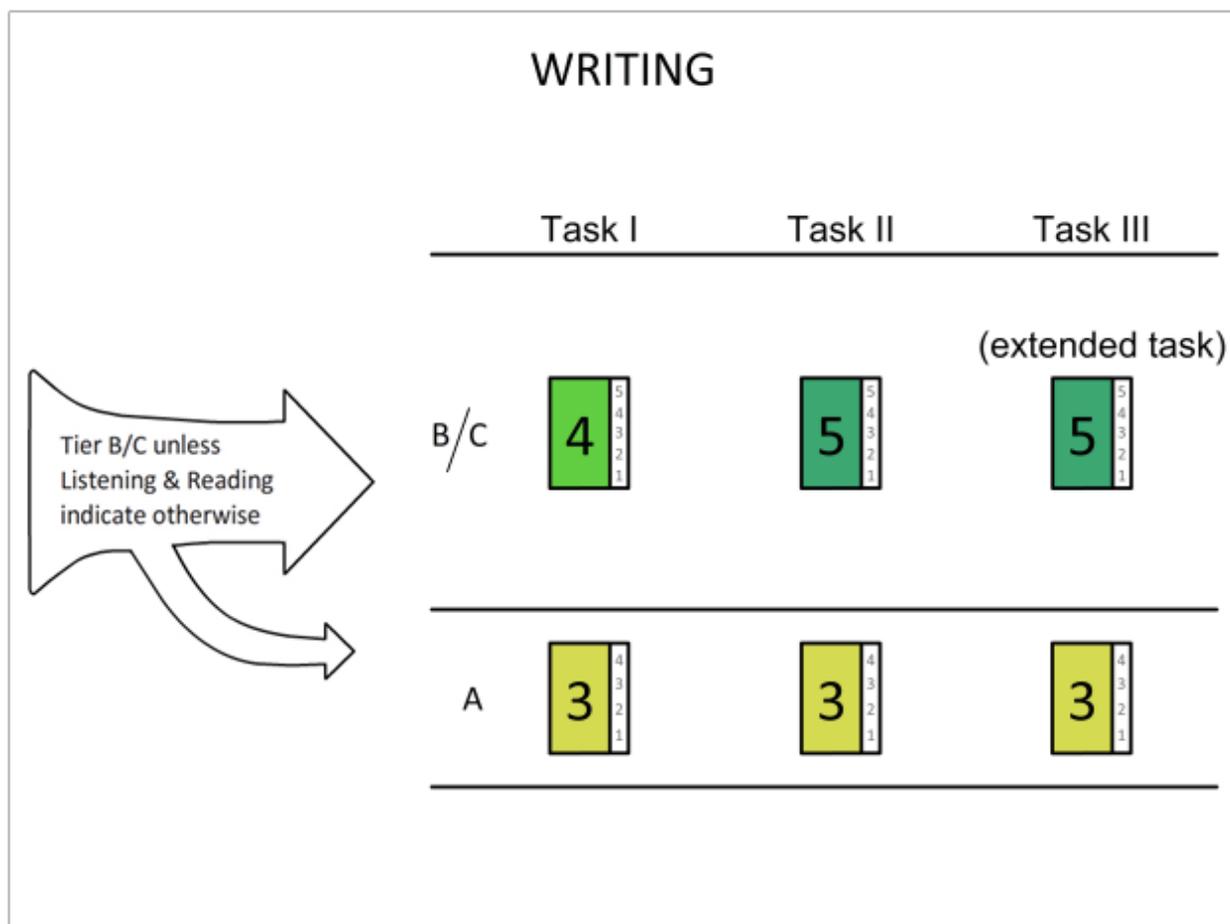
7.2 Tier placement

Tier placement on the Writing subtest will work differently depending on whether students are taking a paper-based or computer-delivered test. All examinees in grades 1-3 will take the Writing subtest on paper, grouped by tier, as will any older student judged to be better served by the paper-based accommodation for the entire test. Tiered paper-based tests will have to be ordered in advance, with tiering decisions made as they are for the current ACCESS. Tier placement for the computer-based Writing subtest, however, will be informed by student performance on the Listening and Reading subtests. While most test takers will be routed by default into the Tier B/C Writing subtest, lesser numbers of test takers may be placed in Tier A on the basis of their scores in the receptive domains of Listening and Reading. Data from the current operational ACCESS program will help inform when it is appropriate to route a student to a tier other than B/C.

Once a test taker is placed in a tier, s/he remains in that tier for the duration of the subtest.

While the bold numbers within the colored rectangles indicate, as in the Listening and Reading diagrams, the MPIs the tasks are written to target, the small gray numbers along the right of each rectangle indicate the range of scores attainable on a given task. These serve as a reminder that though a test taker is routed into and stays within a tier, there is really no cap on the English language proficiency s/he can demonstrate within it.

Figure 17. The structure of the Writing subtest



7.3 Task presentation

Except for examinees in grades 1-3 and those taking the paper-based accommodation, writing prompts will appear on the computer screen. While it may not be practicable to incorporate animations or video into the initial version of ACCESS 2.0, the computer's capability to bring graphics to life may be used on future iterations of the test to increase the chances of eliciting rich writing samples from students. In the spirit of providing test takers with maximal support and making every provision to ensure that they have the opportunity to demonstrate the full extent of their English language proficiency, some modeling may be used to help make task expectations as clear as possible, particularly for test takers currently at the lowest levels of English language proficiency. The first of a series of questions might be filled in already, for example, or a sentence starter might be provided.

7.4 Domain-specific design elements

For the same reason that passages on the Reading subtest will be presented in two columns, the text boxes on the Writing subtest of ACCESS 2.0 will occupy approximately half of the width of the screen.

This adheres to the ACCESS 2.0 principle of keeping each item or task restricted to a single screen. Figure 18 shows a possible look for the Writing subtest.

Figure 18. A possible look for the Writing subtest

The interface consists of a blue header bar with the text "Theme Title". Below the header is a large light blue box labeled "Item Graphic". To the right of the "Item Graphic" box is a white box with a blue border. At the top of this white box is a small blue square with the number "3" inside, followed by the text "Task statement/question". Below this is a large white rectangular area with a thin black border, containing the text "Type here". At the bottom right of the entire interface is a small blue square button with a white right-pointing triangle.

7.5 Student response

Examinees in grades 1-3 and those taking the paper-based accommodation for the entire test will plan and write their responses in paper test booklets. Before they actually begin to write, test takers in Tier B/C will be given scrap paper to use if they choose. Writing on this paper is not scored, and the paper will be destroyed after test administration.

As noted in section 3, some students in grades 4-5 and most students in grades 6-12 taking the computerized version of ACCESS 2.0 will keyboard their responses, although some may write them by hand. Students responding on paper will have been pre-identified and will write their responses in pre-ordered and pre-labeled writing response booklets. A mechanism will be built into the Writing test to ensure that it is clear to the centralized scoring center where—on the computer or in a test booklet—to find each test taker's response.

To determine what editing tools should be made available to students who compose their Writing responses on the computer, ASSETS staff will monitor the work of the content consortia Partnership for the Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium (SBAC) and follow their lead. If students will have such word processing functionality as cut-and-paste and block deleting at their disposal on content tests, ACCESS 2.0 should provide these capabilities as well.

Table 6 summarizes the delivery modes used for various components of the Writing subtest.

Table 6. Delivery modes on the Writing subtest

Grade range	Prompt presentation	Scripting	Planning space	Student response
1-3	paper test booklet	live TA	in test booklet	paper test booklet only
4-5	onscreen	virtual and live TA	scrap paper	writing response booklet (or computer for pre-identified students)
6-12	onscreen	virtual and live TA	scrap paper	computer (or writing response booklet for pre-identified students)

7.6 Scoring

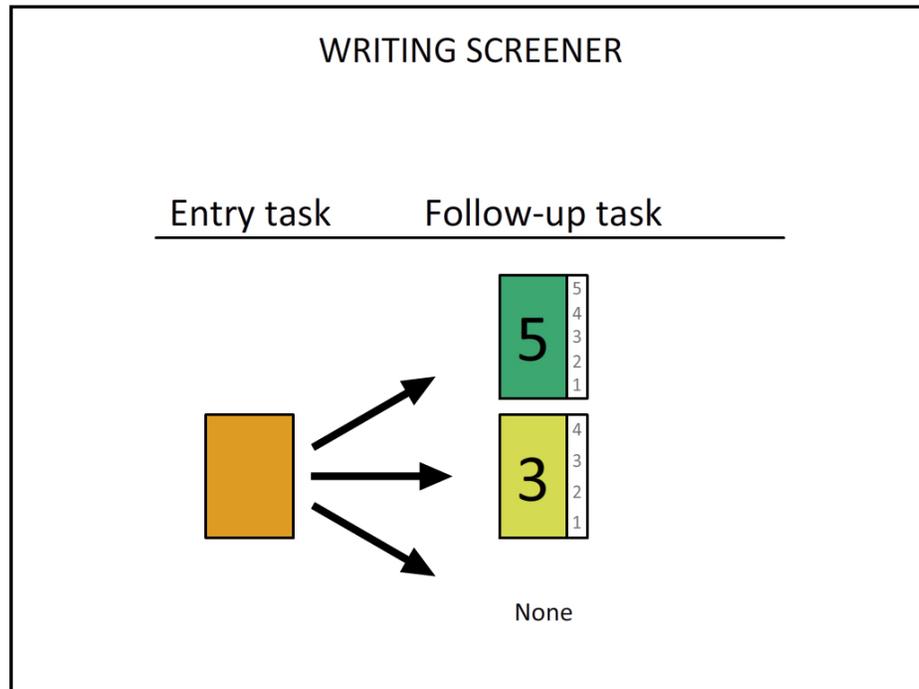
Whether they are keyboarded directly into the computer or written on paper and then scanned, responses to ACCESS 2.0 Writing tasks will be centrally scored by trained raters. For the field testing during the grant period, student writing will be sent to MetriTech, Inc., where highly-trained raters are already familiar with the WIDA rubric for scoring Writing. The Writing rubric currently used for ACCESS for ELLs® will be updated to reflect the 2012 amplification of the WIDA ELD Standards and performance definitions.

7.7 Writing screener

7.7.1 Composition of the Writing screener

In order to get a reliable read on a student’s writing ability, it is necessary to elicit a sizable writing sample. The Writing screener, therefore, will (for test takers with more English language proficiency) include an extended task. In the interests of keeping the screener short, though, this will be preceded by only one other task, as shown in Figure 19. The “entry task,” written to its own specifications, will serve as a check on the test taker’s preparedness for the task to follow. If it becomes apparent during the entry task that a student cannot write a sentence, for example, he or she will exit the screener subsection without being administered a follow-up task. Test takers who have no trouble completing the entry task, however, will be routed to an appropriate follow-up task based on their performance on the Reading and Listening subsections of the ACCESS 2.0 Screener.

Figure 19. The structure of the Writing screener



7.7.2 Scoring the Writing screener

The Writing portion of the ACCESS 2.0 Screener will be locally scored. Materials will be developed as part of the ASSETS grant to guide the local training of raters, whether teachers, administrators, or external hires. A multimedia scorer training program will provide trainee-raters with ample opportunities to practice and to calibrate themselves to a common scale. Specifically, the training program will incorporate the following: (a) interactive, technologically supported training; (b) content-based quizzes that provide diagnostic feedback as a rater learns material; (c) scoring quizzes that provide full justifications for scores and help explain to trainees how to re-calibrate to the scale; (d) computer-adaptive scoring practice that focuses rater-trainees' attention on aspects of scoring they find difficult, using an underlying pool of pre-rated samples, helping raters internalize the criteria; and (e) a digitized library of pre-rated responses that can be used as benchmarks against which to compare writing performances.

Local scoring of written responses on the ACCESS 2.0 Screener will be facilitated through a computerized interface called the ACCESS 2.0 Screener Rater Training Program—Writing. This interface provides the rater with ready access to a test taker's written responses (if digitized). Raters can also view the task instructions, sample scored responses (benchmarks), and tips on scoring, and can easily input ratings and comments.

8 Subtest Specifics: Speaking

8.1 Composition of the Speaking subtest

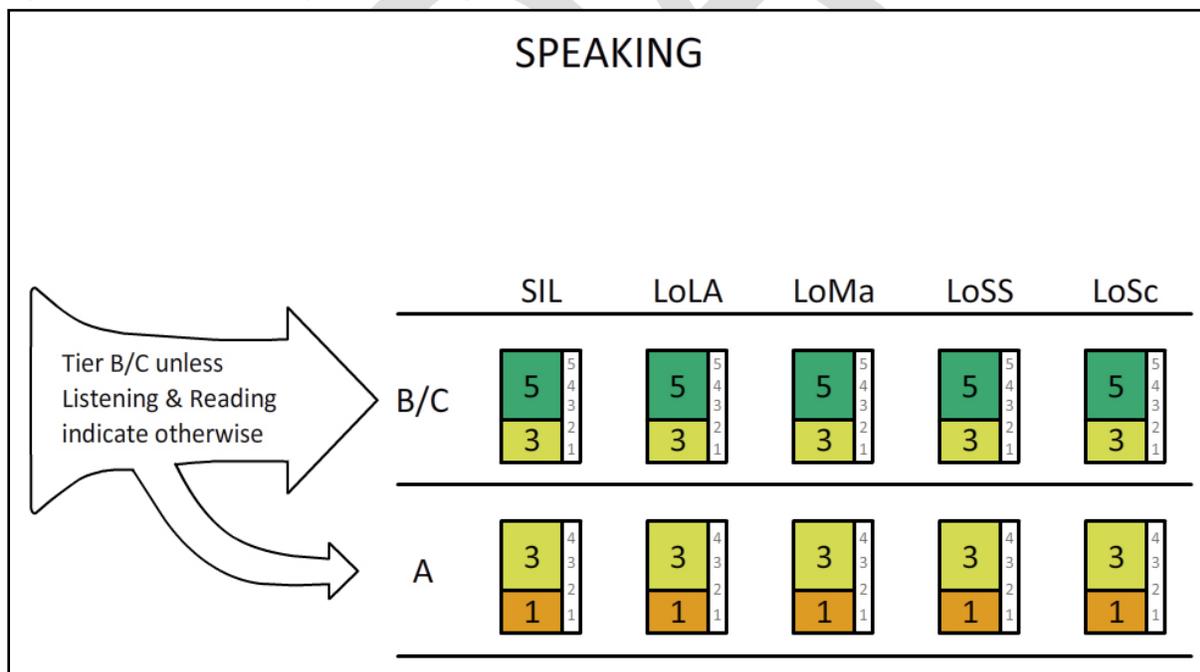
The Speaking subtest of ACCESS 2.0 will comprise five folders, one for each standard. (The decision to move to one folder per standard—Speaking tasks targeted *combinations* of standards on ACCESS for ELLs®—is motivated by the desire to report subscores by standard.) Each folder will consist of two tasks related to a common theme. All students, then, will complete a total of 10 tasks.

8.2 Tier placement

As with Writing, tier placement for the Speaking subtest will be informed by student performance on the Listening and Reading subtests, as indicated in Figure 20. Most test takers will be administered the Tier B/C test, though a student may be routed into Tier A if his or her performance on the Listening and Reading subtests indicates that this is warranted.

Although a student stays within the assigned tier for the duration of the subtest, it is important to note that confining a test taker to a particular tier is not restrictive. The tiers overlap, and tasks are constructed such that students can demonstrate higher than the targeted level of language proficiency. As in the Writing diagram, the small gray numbers along the right of the rectangles in Figure 20 indicate the scores attainable on a given task.

Figure 20. The structure of the Speaking subtest



8.3 Task presentation and student response

As shown in Figure 20, each folder on the Speaking subtest will contain two tasks. Each task will follow the same basic structure and will be designed to support and scaffold student responses using the computer interface. Because it is important for students to speak to an audience, a virtual test

administrator (virtual TA) will guide students through the Speaking tasks. A photograph of the virtual TA will be visible on the screen. The virtual TA will serve as both a narrator and as a virtual interlocutor for the student. A photograph of a model student will also appear on the screen throughout the Speaking subtest. The model student will respond to questions from the virtual TA and demonstrate the language expectations of tasks. Both the virtual TA and the model student will be shown wearing computer headsets.

As students enter a Speaking folder, they will first see a folder introduction screen. The purpose of this screen is to orient the student to the topic and contents of the folder. Both tasks within a folder will be based on the same theme. The folder introduction screen will include a theme graphic and a short description of the general theme, which the student will both hear narrated by the virtual TA and see in text on the screen.

After the folder introduction, students will begin a task. Each task includes three main parts: task input, a model response, and the student response. Table 7 describes the parts of a Speaking task.

Table 7. Parts of a Speaking task

Part	Description
task input	audio, text and graphic support that introduces a student to the topic of the task and provides input about the task theme
model response	model student response to a model prompt that is the same or similar to the task prompt
student response	student response to the task prompt

Within a task, students will first see task input. The task input will include a task graphic (if applicable) and information about the task topic, which the student will hear narrated by the virtual TA and see in text on the screen. If present, the task graphic will remain on the screen while the student responds.

After the introduction screen, the student will hear a model student response. The purpose of the model response is to demonstrate to students the type of response (e.g., quantity of language and level of detail) that is expected. As stated earlier, a photograph of a student wearing a computer headset will appear on the screen throughout the Speaking subtest to represent the model student. The student will hear the virtual TA administer the model prompt (which will be the same as or similar to the task prompt to which the student will later respond). Next, the student will hear a model response. The model response will only be presented by audio and will not appear in text. In cases where the model prompt is identical to the task prompt to which the test taker will respond, the model response will be presented so as not to inhibit test takers' ability to provide an original response. At times, the test taker may hear only the beginning of the model response.

In the third part of a Speaking task, the student will respond to a task prompt. In some cases, the student may hear additional task input after the model student responds to a sample task prompt. The student will use an audio recording mechanism (described in section 8.4) to record responses to a Speaking task. The audio recording mechanism is a part of the test delivery interface. Instructions

reminding the student how to record answers will appear on the screen. After hearing a task prompt, students may record their answers at any time.

Students will then respond to the task prompt. A speech indicator on the audio panel will indicate to students that the recording is working. The expected length of a student response will vary by task level, with higher level tasks requiring more extended discourse (and thus more time to produce). Students may press the stop button when they are done speaking; otherwise the program will automatically stop recording after a certain amount of time has elapsed. Students will then move on to the next task or folder.

8.4 Domain-specific design elements

During the Speaking subtest, students will see, listen to, and respond to tasks via the computer test interface. It is important to note that the Speaking test requires examinees to have a headset with a microphone. From a programming perspective, the Speaking test also necessitates the use of audio capture and audio file storage functionalities.

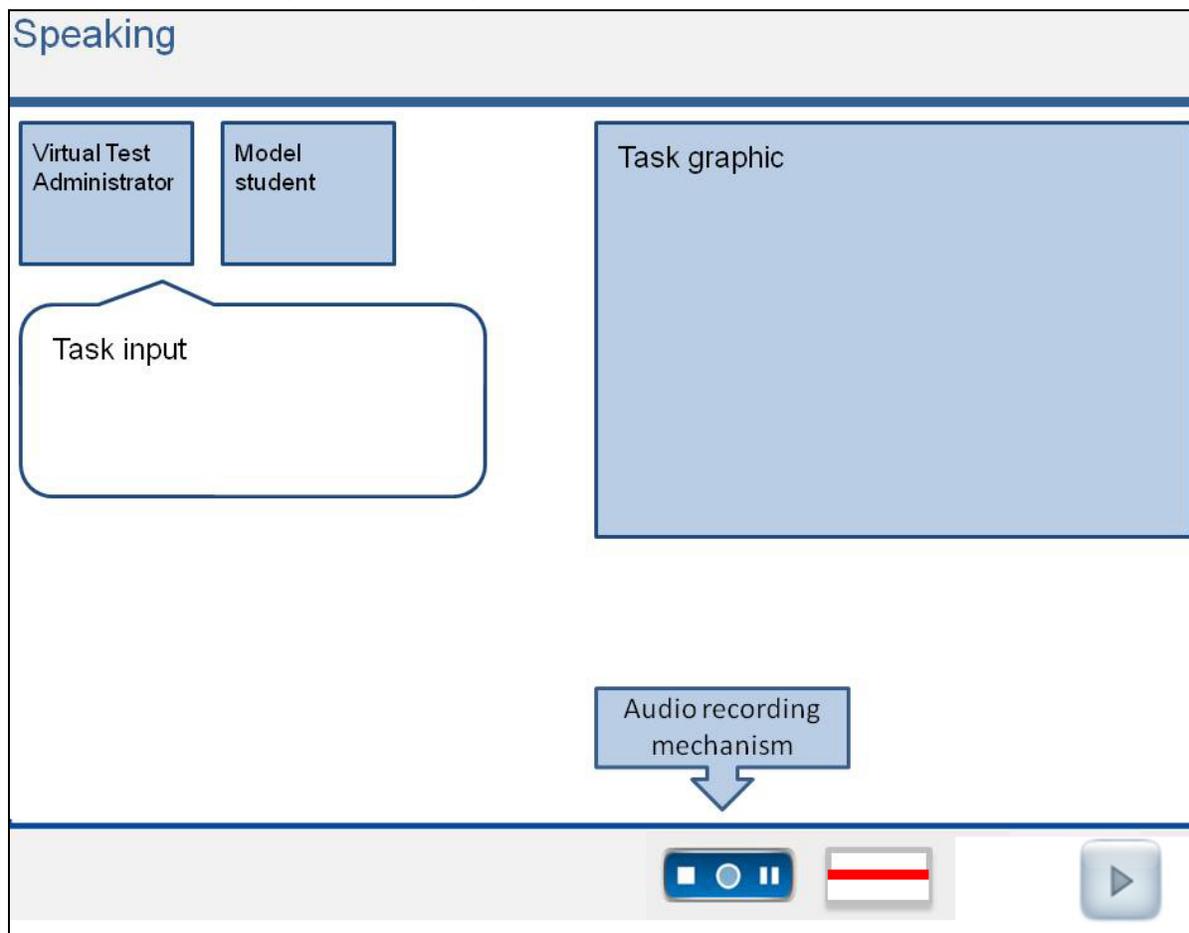
The Speaking subtest will require several domain-specific elements to address the technological demands of this section (i.e., microphone and audio recording) and to facilitate speaking into a computer. A mechanism for students to record their responses is the primary design element that will be unique to the Speaking subtest. This audio recording mechanism is part of the computer delivery interface. While the design of features may vary across computer delivery platforms, certain features are necessary in order for the Speaking subtest to function properly. Table 8 summarizes required features of the audio control panel.

Table 8. Speaking-specific design elements

DESCRIPTION		
The audio recording mechanism is a fixed part of the computer delivery interface and will appear in the navigation panel throughout the Speaking subtest.		
recording mechanism	record/stop button	The student will click a button to begin recording. After the student is done speaking, the student will click a button to stop the recording.
	speech indicator	An onscreen indicator will let students know that the microphone is registering their speech and that it is being recorded.

Figure 21 shows a possible look for the Speaking subtest, incorporating both the audio control panel and a representation of the virtual test administrator.

Figure 21. A possible look for the Speaking subtest



8.5 Scoring

Speaking test responses will be centrally scored. During administration, responses will be captured, saved electronically, and uploaded to the secure scoring system. For the field testing during the grant period, the digitized student oral responses will be sent to MetriTech, Inc.

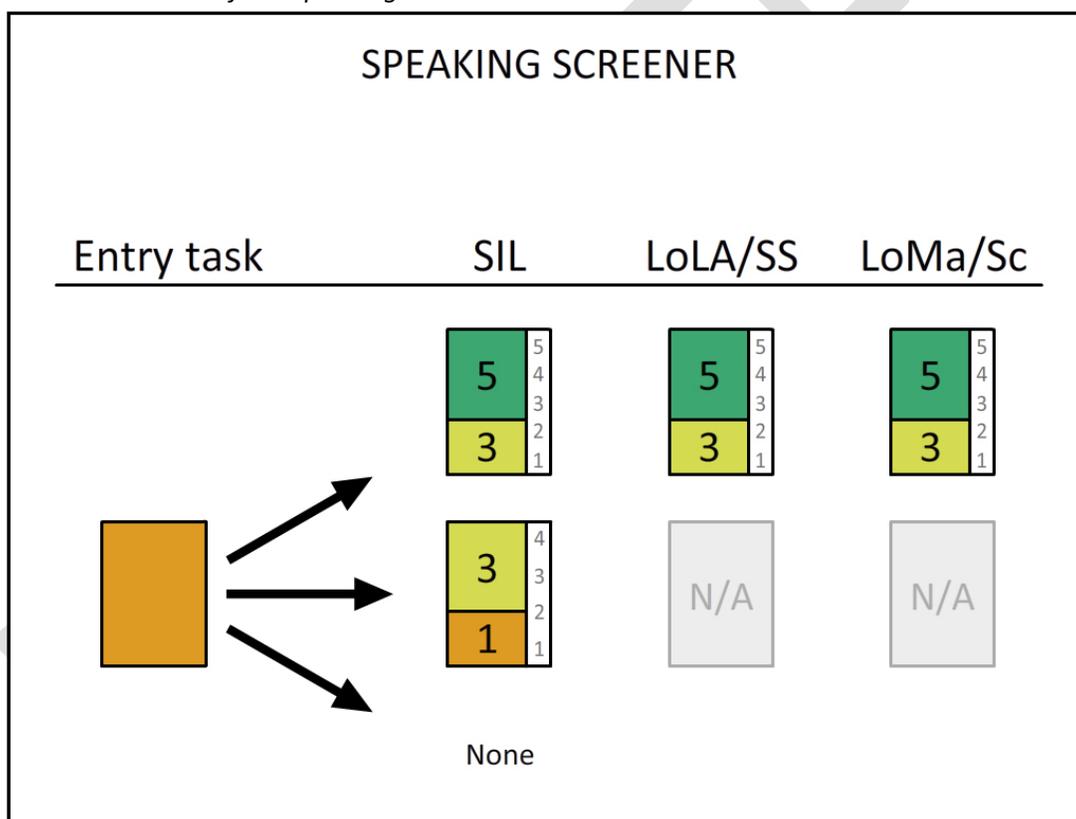
ACCESS 2.0 test development includes the creation of a new rubric for the Speaking test. This rubric will be modeled on the existing ACCESS Writing rubric; in other words, rather than being scored dichotomously as to whether a response meets or approaches task level expectations (as in the current ACCESS), each student response will be given separate leveled ratings on each of the three performance criteria of the WIDA standards: language complexity at the discourse level, language forms and conventions, and vocabulary usage. Although tasks are targeted to elicit language at a particular proficiency level, responses may be scored across the range of proficiency levels represented on the rubric. For example, a rater might indicate that vocabulary use on a proficiency level 3 task was characteristic of vocabulary use at level 4. In this way, the rubric and scoring procedure will allow for the collection of more nuanced information about oral performance.

8.6 Speaking screener

8.6.1 Composition of the Speaking screener

As with the Writing screener (see 7.7.1), there will be three routes through the Speaking screener. An entry task will be used to check a test taker's readiness for additional tasks. A student demonstrating no English language proficiency in speaking on the entry task will complete only that task, while other test takers will be administered either one or three follow-up tasks, depending on their performance on the Reading and Listening subsections of the ACCESS 2.0 Screener. For a student demonstrating a high level of English language proficiency, the Speaking screener will consist of three folders, one for SIL, one for either LoLA or LoSS, and one for either LoMa or LoSc, preceded by an entry task. Figure 22 shows the structure of the Speaking screener.

Figure 22. The structure of the Speaking screener



8.6.2 Scoring the Speaking screener

The Speaking portion of the ACCESS 2.0 Screener will be locally scored by educators trained using a computer-based scorer training program called the ACCESS 2.0 Screener Rater Training Program—Speaking, based on the Multimedia Rater Training Program (MRTP) developed by the Center for Applied Linguistics. The scorer training will incorporate the following: (a) interactive, technologically supported training, (b) content-based quizzes that provide diagnostic feedback as a rater learns material, (c) scoring quizzes that provide full justifications for scores, (d) computer-adaptive scoring practice (using an underlying pool of pre-rated samples) that focuses rater-trainees' attention on aspects of scoring

they find difficult and helps rater-trainees internalize the scoring criteria, and (e) a digitized library of pre-rated responses that can be used as benchmarks against which to compare student performances.

In addition, local rating of oral responses on the ACCESS 2.0 Screener will be facilitated through a computerized interface adapted from the Center for Applied Linguistics' Computerized Oral Proficiency Instrument. This interface will provide the rater with ready access to a test-taker's oral responses, with full playback/pause/repeat functionality. Raters can also view the task instructions, sample rated responses (benchmarks), and tips on rating, and can easily input ratings and comments.

9 Next Steps

The *Test and Item Design Plan* is not an end in itself, but a means toward the end of getting ACCESS 2.0 and the ACCESS 2.0 Screener operational by 2015-16. Table 9 serves as a reminder of the phases of development yet to come. This document has informed and been informed by the task and item prototyping currently underway in preparation for the pilot testing phase of the ASSETS project.

Table 9. Phases of development

Year	Phase of Development	Project Activities
2011-12	Initial development	<ul style="list-style-type: none"> • create initial overall test design and development plan and initial test and item specifications, with input and consensus from stakeholders • develop and review item pool for piloting and assemble pilot-test forms • develop ancillary materials for pilot test (including administrator/scorer materials)
2012-13	Pilot testing	<ul style="list-style-type: none"> • pilot items through iterative cognitive labs and tryouts • conduct analyses on data collected through cognitive labs and tryouts • revise overall test design and development plan, including generative test and item specifications (as needed)
2013-14	Field testing	<ul style="list-style-type: none"> • develop and review item pool for field test • assemble field-test forms • develop ancillary materials for field test (including administrator/scorer materials) • conduct field test • score field test • conduct field-test item analyses and reliability studies • conduct field-test validation studies • finalize design of all components of operational testing program

2014-15	Pre-operationalization	<ul style="list-style-type: none"> • conduct standard setting • assemble final operational test forms • develop final score reports and score reporting system • develop ancillary materials for operational test • finalize plan for training and monitoring administrators/scorers and assemble final training materials • conduct additional reliability studies • conduct additional validation studies • finalize plan for continual monitoring and evaluation of operational testing program
---------	------------------------	--

10 Appendix: Understanding the ACCESS for ELLs® Test

10.1 Introduction

As mentioned in sections 1.2 and 2.4, this appendix has been included with the *Test and Item Design Plan* to give readers unfamiliar with ACCESS for ELLs® background about ACCESS 2.0’s illustrious precursor.

ACCESS for ELLs (Assessing Comprehension and Communication in English State-to-State for English Language Learners) is a secure large-scale English language proficiency assessment given to students in grades K-12² who have been identified as English language learners (ELLs). It is given annually in WIDA Consortium member states to monitor students’ progress in acquiring academic English. This document explains the structure of the test, including the grade-level clusters and tiers.

10.2 Test structure

The ACCESS for ELLs test battery is a collection of assessment instruments administered to all ELL students across grades K-12. It first and foremost operationalizes the **English language development standards**³ that form the core of the WIDA Consortium’s approach to serving the needs of English language learners. These standards incorporate a set of **model performance indicators (MPIs)** that describe the academic language proficiency expectations for ELL students at five different grade-level clusters and in five different content areas.

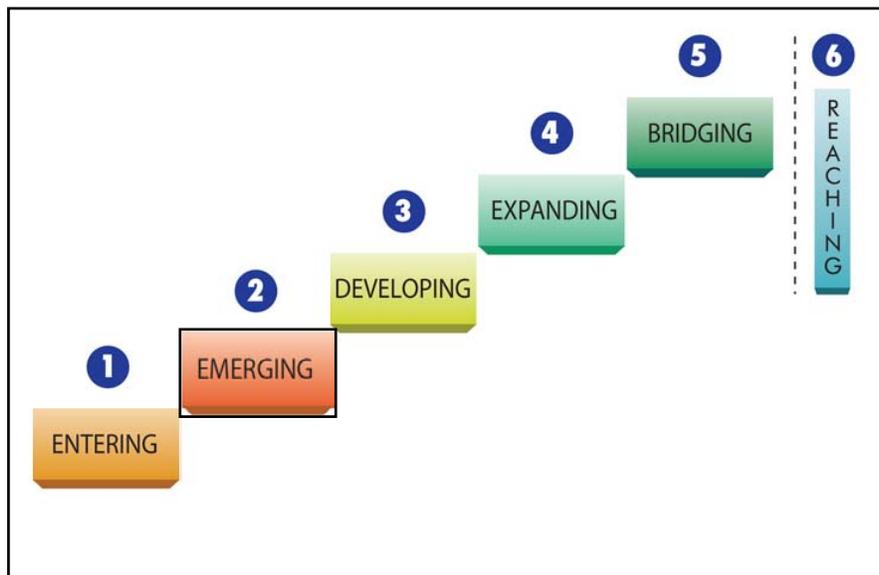
The **grade level clusters** of the standards are Pre-K–K, 1–2, 3–5, 6–8, and 9–12. Test forms follow this grade-level clustering as well, though the K test is appropriate only for kindergarten, and not for pre-kindergarten. The WIDA framework recognizes the continuum of language development within the four language domains across six proficiency levels.

² Remember that while ACCESS for ELLs® includes a kindergarten test, this *Test and Item Design Plan* is concerned with the summative test and screener for **grades 1-12 only**. The development of a new kindergarten test is not funded under the ASSETS grant, nor is the development of the Alternate ACCESS for ELLs.

³ In 2012, WIDA recharacterized its English Language Proficiency Standards as English Language Development Standards to emphasize that the process of language development is fluid and flexible.

The six levels depicted in Figure 23 describe the spectrum of a learner’s progression from knowing little to no English to acquiring the English skills necessary to be successful in an English-only mainstream classroom.

Figure 23. The six proficiency levels



The ACCESS for ELLs test incorporates items assessing academic language from each of the **five content areas** of the standards. The first of these is Social and Instructional language (SIL), which incorporates proficiencies needed to deal with the general language of the classroom and the school. The other standards are the language of English Language Arts (LoLA), the language of Mathematics (LoMa), the language of Science (LoSc), and the language of Social Studies (LoSS).

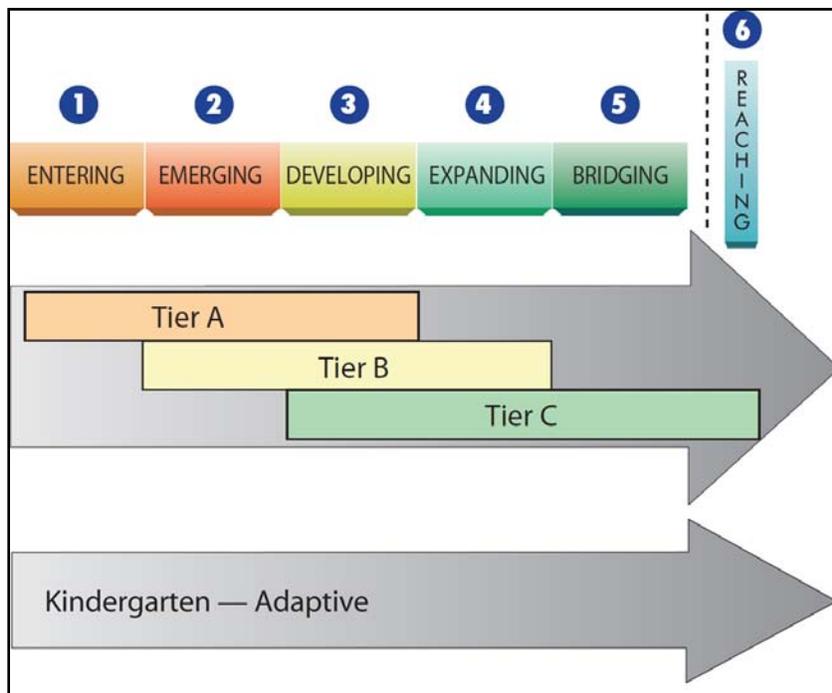
For each grade-level cluster, the standards specify one or more model performance indicators for each content area within each of the **four language domains**: Listening, Speaking, Reading, and Writing. Test items are written to specifications that address particular performance indicators. Each performance indicator and the items developed from it address one of five proficiency levels which array themselves to describe a developmental continuum of language proficiency.

10.3 Tiers

The goal of the ACCESS for ELLs test is to allow students to demonstrate their level of proficiency through the MPIs. However, a test with questions assessing each and every MPI would be far too long to fit in any reasonable testing session. Too many easy questions make a test boring, while too many hard ones make it frustrating. It is important to avoid both possibilities to achieve a reliable test. To ensure that the test an individual student takes is appropriate for his or her incoming English language

proficiency, test items are presented in 3 **tiers** for each grade-level cluster; these are simply designated as Tiers A, B, and C. Figure 24 shows how the different tiers map to the proficiency levels.

Figure 24.



As illustrated in Figure 24, the tiers overlap: Tier A covers levels 1-3, Tier B covers levels 2-4, and Tier C covers levels 3-5. The Tier C test also includes some items that are slightly more difficult than proficiency level 5. The kindergarten test is not tiered.

Each tier, of course, is only able to discriminate performance on its portion of the proficiency scale, so to make sure that the whole ACCESS for ELLs test works as intended, it is necessary to place each student into the tier that best matches his or her proficiency level. The decision as to where the student currently falls on the scale is best made by the student’s teachers, based on the information they have about the student’s language proficiency, including performance on some other language tests, such as the WIDA-ACCESS Placement Test (W-APT™).

For students lacking other test scores, the WIDA Consortium provides a protocol—shown in Figure 25—for placing students into the appropriate tier of the test. Matching students to tiers appropriately will maximize accuracy and validity of the results. Note that English language learners must meet at least ONE of the criteria listed under each tier to qualify for that tier.

Figure 25.

Level 1 Entering	Level 2 Emerging	Level 3 Developing	Level 4 Expanding	Level 5 Bridging	Level 6 - Reaching
<p>TIER A is most appropriate for English language learners who:</p> <ul style="list-style-type: none"> • have arrived in the U.S. or entered school in the U.S. within this academic school year without previous instruction in English, OR • currently receive literacy instruction ONLY in their native language, OR • have recently tested at the lowest level of English language proficiency 					
<p>TIER B is most appropriate for English language learners who:</p> <ul style="list-style-type: none"> • have social language proficiency and some, but not extensive, academic language proficiency in English, OR • have acquired some literacy in English though have not yet reached grade level literacy 					
		<p>TIER C is most appropriate for English language learners who:</p> <ul style="list-style-type: none"> • are approaching grade level in literacy and academic language proficiency in the core content areas, OR • will likely meet the state's exit criteria for support services by the end of the academic year 			

10.4 Test administration times

The target administration times for each section of the test are:

- Listening: 25 minutes, machine scored
- Reading: 35 minutes, machine scored
- Writing: up to 1 hour, rater scored
- Speaking: up to 15 minutes, administrator scored

The group-administered tests at grade levels 1-12 are given in two 75-minute sessions: Session 1 for the Listening and Reading tests and Session 2 for the Writing test. The Speaking test must be administered to students individually in a separate third session.

The entire kindergarten test is administered individually, and, depending on the student's proficiency, may take up to 45 minutes to administer. More detailed instructions for administering ACCESS for ELLs are contained in the test administration manuals.