ALTELLA

ALTERNATE ENGLISH LANGUAGE LEARNING ASSESSMENT PROJECT

Developing Item Templates for Alternate Assessments of English Language Proficiency

ALTELLA Report

August 2018

Phoebe C. Winter, Meagan Karvonen, and Laurene L. Christensen

Wisconsin Center for Education Research
UNIVERSITY OF WISCONSIN–MADISON

Developing Item Templates for Alternate Assessments of English Language Proficiency

ALTELLA Report August 2018

Phoebe C. Winter, Meagan Karvonen, and Laurene L. Christensen

Suggested Citation

Winter, P. C., Karvonen, M., & Christensen, L. L. (2018, August). *Developing item templates for alternate assessments of English language proficiency.* Madison, WI: University of Wisconsin–Madison, Alternate English Language Learning Assessment (ALTELLA). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research: http://altella.wceruw.org/resources.html

Updated: 10/3/2018

The Alternate English Language Learning Assessment (ALTELLA) project researches instructional practices, accessibility features and accommodations, and assessment of English learners with significant cognitive disabilities to develop an evidence-centered design approach that informs our understanding of alternate English language proficiency assessment for these students.

ALTELLA is a partnership of five state departments of education and the Wisconsin Center for Education Research at the University of Wisconsin–Madison. This collaboration of states involving Arizona, Michigan, Minnesota, South Carolina, and West Virginia is funded by an Enhanced Assessment Instruments grant from the U.S. Department of Education awarded to the Arizona Department of Education. ALTELLA is housed within the Wisconsin Center for Education Research.

altella.wceruw.org

ALTELLA Staff

Laurene Christensen, principal investigator, laurene.christensen@wisc.edu

Erin Arango-Escalante
Kristen Burton
Indira Ceylan
Elizabeth Cranley
Melissa Gholson
Jonathan Gibson
Miguel Hernandez
James Mitchell
Sarah Ryan
Kristopher Stewart
Sonia Upton
Cha Kai Yang

Publication Coordinator Indira Ceylan
Report Editor Karen Faster
Graphic Designer Janet Trembley

ALTELLA Project Item Templates Activity Participants

Audra Ahumada, Arizona Department of Education
Cheryl Alcaya, Minnesota Department of Education
Jill Christmus, South Carolina Department of Education
Mami Itamochi, West Virginia Department of Education
John Jacquith, Michigan Department of Education

Table of Contents

List of Figures

Introduction

This report provides an exploration of how item and task development, a critical component of assessment design, might be implemented for an alternate assessment of English language proficiency standards), keeping at the forefront the population of students taking the test and the ways English language proficiency might be defined for this population. The quality of inferences made from assessment results depends in large part on the quality of test components—the relationships between the tasks and the test claims and purposes, the nature of the tasks the students encounter, the interaction between the task and student, and the ways the tasks are scored.

The Alternate English Language Learning Assessment (ALTELLA) project convened a group of state department of education staff with expertise in educating English learners and students with significant cognitive disabilities and staff with expertise in large-scale assessment design to discuss considerations for item templates that could be used in the development of an alternate assessment of English language proficiency. This document grew out of that meeting and is intended to highlight considerations in developing test items and tasks to measure the language proficiency of students who are classified as English learners with significant cognitive disabilities. In line with the panel members' perspectives, the report is based on the principle that an alternate assessment of English language proficiency should be maximally accessible in order to be appropriate for measuring the language skills of this population.

Background

The U.S. Elementary and Secondary Education Act, as amended by the Every Student Succeeds Act (2015), requires state education agencies to provide for the annual assessments of the English proficiency of all students identified as English learners (Section 3111(b)(2)(G)). This includes students with significant cognitive disabilities who are English learners. In October 2017, the U.S. Department of Education made it clear that all students must be assessed, issuing a memo that referenced 34 CFR § 200.6(h)(5). This regulation requires states to provide an alternate English language proficiency assessment for English learners with the most significant cognitive disabilities, students who cannot participate in the general English language proficiency assessment even with appropriate accommodations. The memo notes that states may develop alternate English language proficiency achievement standards for this alternate assessment (Rooney, 2017).

State departments of education have long recognized the need to be able to ascertain whether students with significant cognitive disabilities are also English learners. English

learners with significant cognitive disabilities need to receive appropriate support services and assistance in learning English. However, until the 2017 memo from the U.S. Department of Education, there were no provisions for developing alternate English language proficiency standards. Potential English learners with significant cognitive disabilities had to be assessed against the same achievement standards used for the state's general test. States that develop or adopt alternate English language proficiency standards[1] will need assessments designed to measure these standards. These assessments must meet the technical quality guidelines specified by the U.S. Department of Education's peer review process (see 34 CFR § 200.2(b)(2), (4), and (5)).

English Learners with Significant Cognitive Disabilities

Christensen, Gholson, and Shyyan (2018) define English learners with significant cognitive disabilities as students "who have one or more disabilities that significantly limit their intellectual functioning and adaptive behavior as documented in their Individualized Education Programs, and who are progressing toward English language proficiency in speaking, reading, writing, and understanding" (p. 3). States will need guidelines for local education agencies to use in identifying students as English learners with significant cognitive disabilities eligible to take an alternate assessment of English language proficiency. At a minimum, the primary language spoken in the student's home is likely to be other than English, and the student will meet criteria for taking an alternate assessment developed by the state for its academic content assessments, if the student is in a tested grade. Appropriate identification is complicated by the fact that states are responsible for identifying and assessing English learners beginning in kindergarten, sometimes before significant disabilities are identified (Christensen et al., 2018). In other cases, identifying the student's status as an English learner may be challenging because the assessments used for screening students may not be optimally accessible to these students with disabilities.

Little research has explored the characteristics and learning needs of English learners who have significant cognitive disabilities (Shyyan, Gholson, & Christensen, 2018; Thurlow, Christensen, & Shyyan, 2016). For example, students with significant cognitive disabilities who are English learners tend to have less complex expressive communication and lower performance on English language arts assessments than their non-English learner counterparts (Clark, Karvonen, Wells-Moreaux, & Erickson, 2018). ALTELLA project

[1] English language proficiency standards address the development of communication skills in listening, speaking, reading, and writing, with the goal of students developing the language skills needed to acquire and demonstrate the knowledge and skills in state academic content standards. Alternate English language proficiency standards will need to describe expectations related to how communication skills develop in students with significant cognitive disabilities. However, there is no federal definition of alternate English language proficiency standards.

researchers collected survey data about characteristics of English learners with significant cognitive disabilities. The 29 states that participated in the study report the following about their students whose home language was not English and who had significant cognitive disabilities (Christensen, Mitchell, Shyyan, & Ryan, 2018):

- Seventy-one home languages were found, with Spanish (53.8%), English (27.8%), and Arabic (2.6%) being the top three languages.
- Primary disabilities break down as follows: 42.3% intellectual disability; 11.6% multiple disabilities; 26.9% autism; and 6.0% developmental delay.
- Students who had not taken an English language proficiency assessment made up 31.2%.
- Most of these students receive their content instruction in self-contained classrooms for students with disabilities while their English development support includes full bilingual instruction, push-in models, pull-out models, consultative models, and no English development support (Christensen & Mitchell, 2018).

Developing Item Templates for an Alternate English Language Proficiency Assessment

Assessment items and tasks should be designed to elicit information relative to test claims, which themselves should be based on the purposes and uses of assessment results, a close analysis of the domain(s) to be assessed, and on characteristics of the student population. Item templates flow from the evidence-centered design process, and the prerequisites for developing item templates will be based on work done by test designers as they implement evidence-centered design. The contents of this report are not based on the elucidation of specific claims for an alternate English language proficiency assessment or on any other features of the work that needs to be done before developing item templates. Rather, this report focuses on the general components item templates might contain and on the considerations surrounding their development, rather than providing models of item templates for specific assessment contexts and claims.

A primary purpose of an alternate English language proficiency assessment is to ensure that English learners with significant cognitive disabilities can benefit from instruction in classrooms where English is the language of instruction. In these classrooms, students with significant cognitive disabilities, both English learners and students who are not English learners, are afforded multiple ways to communicate in instruction and assessment, dependent on the nature of their disabilities and individual learning needs. The considerations in this report for developing item templates reflect the reality of classroom

instruction and are written from the perspective of maximum accessibility. An alternate English language proficiency assessment should allow for the same modes of communication used by non-English learners in classrooms where English is the mode of instruction.

Item Template Development and Evidence-Centered Design

The idea of creating item templates is an important component of evidence centered design. Evidence centered design is not a set of rigid procedures, but rather a principles-based process for assessment design and development (Hauck, Pooler, & Anderson, May 2015). Mislevy (2011, p. 6) describes evidence centered design evidence-centered design as "an evidentiary argument for reasoning what students say, do, or make in particular task situations as well as to generally claim what they can know, do, or have accomplished." Mislevy and Haertel (2006) describe the five "layers" of evidence centered design:

1. Domain analysis, which consists of specifying the central aspects of the domain to be assessed;
2. Domain modeling, which consists of laying out the claims intended to follow from the assessment – the assessment argument – that stem from the domain analysis;
3. Conceptual assessment framework, which requires specifying (a) the evidence that will support the claims the assessment is intended to make about students, (b) the contexts surrounding the collection of that evidence, and (c) how tasks and items will be designed to collect the evidence needed about student achievement/proficiency;
4. Assessment implementation, which consists of creating the elements specified in the conceptual assessment framework, e.g., developing items and rubrics, creating forms, constructing the delivery system; and
5. Assessment delivery, consisting of students taking the assessments, response scoring, and score reporting.

Item design and development, the focus of this report, is part of the conceptual assessment framework layer, and draws on the domain analysis and domain modeling layers. It is beyond this report's scope to go into detail about other aspects of evidence-centered design. Good resources for understanding more about evidence-centered design and its implementation include Mislevy and Haertel (2006) and Cameto, Haertel, DeBarger, and Morrison (2010).

Item Template Development and Universal Design for Learning

Universal design for learning is a set of principles for designing assessments from the beginning so that they are accessible for a wide range of students, regardless of learning needs or disabilities. While universal design for learning is not an explicit part of evidence-centered design, using universal design for learning principles to develop items is a logical component of the evidence-centered design process. These universal principles help define the context and form of the evidence that supports the assessment's claims, as specified in layer 3, the conceptual assessment framework, of the evidence-centered design process.

The Center for Applied Special Technology (2018) categorizes universal design for learning guidelines into three categories: providing multiple means of engagement; providing multiple means of representation; and providing multiple means of action and expression. Researchers at the National Center on Educational Outcomes lay out considerations for incorporating universal design for learning into assessments:

- Intended constructs are measured
- Respect for the diversity of the assessment population
- Concise and readable text
- Clear format for test
- Clear visuals
- Changes allowed to format without changing meaning or difficulty
  (Johnstone, Altman, & Thurlow, 2006, p. 1)

As with evidence-centered design principles, universal design for learning should permeate assessment design. It is beyond the scope of this report to go into detail about universal design for learning. Good resources that further describe and illustrate universal design for learning in an assessment setting are Center for Applied Special Technology (2015, 2018); Christensen, Shyyan, and Johnstone (2014); Johnstone, Altman, Thurlow, and Moore (2006); and Johnstone, Altman and Thurlow (2006). For a discussion of how universal design for learning fits into evidence-centered design, see Gordon, Gravel, and Schifter (2010).

Characteristics of Item Templates for an Alternate English Language Proficiency Assessment

This report, uses the term "item template" to refer to item and task design specifications. Item templates provide a framework for collecting evidence about the claim made about student achievement based on test scores. A template specifies:

- The measurement target. Earlier components of the evidence-centered design process have laid out the evidence needed to support the claims made by the assessment. The measurement target defines which aspects of that evidence the particular item or task addresses.
- Parameters around how the item or task should be constructed, such as contexts, levels of difficulty, and complexity. Earlier components of the evidence-centered design process include considerations of how the measurement target contributes evidence to the claim; the claims, the types of evidence supporting the claims, and the characteristics of the student population will guide these parameters of task development.
- Rules and rubrics for scoring the item or task. The scoring rules indicate the level of skill the student has regarding the measurement target, and the accumulation of item and task scores provides overall evidence for student status on the test's claims.

Examples of Item Templates

Several testing programs have followed evidence-centered design principles in development, using very different-looking "item templates." The Multi-State Alternate Assessment, English Language Proficiency Assessment for the 21st Century, WIDA, and Dynamic Learning Maps illustrate the range of item specifications that can be used in designing assessments with evidence-centered design principles.

The Multi-State Alternate Assessment (formerly the National Center and State Collaborative assessment) consists of content area assessments for students with significant cognitive disabilities. Its templates are very similar to those used by the earlier Principles Assessment Designs for Inquiry project, which followed a type of evidence-centered design (Riconscente, Mislevy, & Hamel, 2005). Multi-State Alternate Assessment templates are designed to be used in an online item development system. Selected variable features are populated based on the conditions of assessment for the task and for each level of the task. (Each Multi-State Alternate Assessment item template was originally created to address four levels of complexity, however, based on new internal studies these have since

been changed to addressing the levels of complexity.) For example, a language arts template might specify the genre of the passage and its length, sentence complexity, and overall complexity level. The template could specify the targeted concept; precursor knowledge, skills, and abilities needed to respond as well as the targeted knowledge, skills, and abilities; specific cognitive and mechanical supports that are either embedded within the task or the administrator could use at each level; different allowable response modes; and the number and characteristics of options the examinee chooses from.

English Language Proficiency Assessment for the 21st Century, a test of English language proficiency, used evidence-centered design to create an assessment that elicits responses providing evidence of language acquisition. The item template design began with conceptualizing how students could show their standing on the English language proficiency standards that are the basis of the test—that is, the student behaviors in language that would provide evidence of language proficiency. English Language Proficiency Assessment for the 21st Century item template design features are specified at three levels: full test, domain, and task type. Accessibility considerations, including principles and procedures for building in universal design, available accessibility tools and accommodations, and considerations for students who are blind or have other vision impairments, are specified at the test level. At the domain level (domains are reading, listening, writing, and speaking), stimulus and response contexts, characteristics, parameters, and limitations are specified. Finally, at the task type level, the standard(s) addressed by the task type and more detailed task characteristics are described, all based on an explication of the types of evidence and scope and boundaries around that evidence needed to address the measurement target. Examples of one or more tasks accompany each template.

WIDA's ACCESS for ELLs 2.0 assessment also used an evidence-centered design approach in the development of its test of English language proficiency. Items and tasks are designed to reflect the elements of the domain analysis needed to collect the necessary evidence required to determine the student's progress toward attainment of proficiency in that domain. Each item in the listening and reading domains are written to reflect the language of one of the five WIDA English Language Development Standards and to test a student's ability to process language at one of the five fully delineated proficiency levels. Writing tasks are designed to elicit language from one or more of the WIDA English Language Development Standards. Task modeling is an essential component of the speaking test design. Students are introduced to a virtual model student and a virtual model administrator during the test directions. Prior to responding to each task, test takers first listen to the model student respond to a parallel task. The purpose of the model is to demonstrate task expectations to both test takers and to raters who score all speaking task responses.

Dynamic Learning Maps, another assessment of academic content achievement for students with the most significant cognitive disabilities, has a more concise template, capitalizing on links to associated documents that further define the measurement target and contextual constraints. Developed for each content standard, Dynamic Learning Maps' Essential Element Concept Maps (EECM) identifies the associated claim and conceptual area for the standard. Because items are written in the context of a testlet that also includes an engagement activity, the EECM specifies the number of items per skill that need to be written for the testlet. The EECM includes descriptions of student behaviors that exemplify performance of the skill, without limiting performance to a certain response mode. For each skill, the EECM lists common student misconceptions along with questions that typically elicit responses that would demonstrate student mastery of the skill. The EECM also lists concepts and vocabulary associated with the skill. Rather than providing long descriptions of non-targeted skills, the EECM contains the same type of information (e.g., descriptions of student behaviors, questions to ask, misconceptions, concepts, vocabulary) for related skills in the learning map. The EECM is accompanied by a graphic of a small section of the learning map so item writers can see the relationships of target skills to non-target but related skills. The EECM also flags potential barriers to assessing the specific skill due to sensory or physical challenges. The flag provides additional cautions beyond the standard Dynamic Learning Maps item writing guidance and may point to the need for an alternate form (e.g., for students with visual impairments) or special administration instructions (e.g., for a student who cannot manipulate objects to demonstrate a process). Other information commonly found in item templates is addressed through ancillary materials. For example, the item writing handbook provides general rules about choice of item types, and item writers are expected to apply those guidelines across testlets.

Considering the Student Population

The range of student characteristics in the targeted assessment population must be taken into account when developing item templates. The population of English learners with significant cognitive disabilities is extremely diverse. Each template needs to address factors that may affect both the content of the task and the administration conditions necessary to elicit student responses that provide evidence related to its measurement target.

Test design and administration plans should take into account that many of the students may have little or no experience in taking English language proficiency assessments or taking assessments in the particular form (e.g., online) that the test is in. Administration plans should include procedures that familiarize students with test content and administration flow, such as providing meaningful practice in situations mimicking those on the test, so that students can demonstrate their proficiency when they take the actual assessment.

Considering the Test Structure and Purpose

Decisions about how English language proficiency is defined and assessed will affect which features are included in an item template and how those features are specified. For example, the state may determine that reading, listening, writing, and speaking should be assessed separately; that receptive language (reading and listening) and expressive language (writing and speaking) should constitute the sections of the assessment; or that a single test section should address all features of language acquisition, using interactive tasks. These decisions will affect the measurement targets and the specifications of various item characteristics (e.g., nature of the stimulus) and what types of accessibility tools and supports are allowed.

Regardless of assessment structure, decisions will need to be made about which forms of communication to include in the alternate English language proficiency assessment for each student. The federal regulations governing English language proficiency assessments allow students to be exempted from a domain or domains if appropriate accommodations cannot be given to the student (CFR 200.6(h)(5)(ii)). However, if the state has determined that the purpose of the test is to determine whether the student can benefit from instruction delivered in English in the classroom and has explicitly defined this instruction as including accessibility conditions available to students who are not ELs, it is quite likely that the need to exempt domains will be rare.

Features of Item Templates

While different assessment programs use different styles of item templates, there are common components across template styles. There are several features that should be considered when developing item templates for an alternate assessment of English language proficiency (Figure 1). These components are derived from a review of item templates and item specifications across several assessment programs, in light of the descriptions of evidence-centered design in papers by Mislevy and colleagues, cited earlier. Some may be explicit parts of each item template, varying by measurement target; others may be more general and apply to a domain, a modality, or the test as whole. All features should be should be based on the definition of language acquisition and how English language proficiency manifests itself in this population, based on outcomes of earlier stages of the evidence-centered design evidence-centered design process. As item templates are being developed, it may become apparent that claims or other aspects of the domain modeling stage need to be revised. Evidence-centered design allows for an iterative process in test design and development so that the final product is internally consistent and coherent.

The effects of student characteristics on administration and response modes should be explicitly addressed and considered throughout the template development process. For

example, an item that in its most typical form may require a student to move an object on a touchscreen must have provisions for students who cannot manipulate objects on a touchscreen. In some cases, items may need different formats to be accessible to some students; for example, a text-based item may need to be rendered in braille or read aloud, or an oral item signed to a student. In some cases, similar items that do not have identical content may need to be developed. For example, some items may depend on pictures for their content and the tactile graphic version may be too complex for students with significant cognitive disabilities, requiring the development of a substitute item measuring the same concept. Other considerations apply to the test as a whole. For example, if colors are used on the test, they should be selected to be appropriate for students with various visual impairments.

The features in the following list of potential template contents are not independent of each other. For example, feature 5, limitations/parameters around measurement target, will affect feature 7, stimulus specifications/limitations.

Figure 1: List of potential features of item templates

Note: The feature numbers correspond to annotations in Figure 2.

1. **Domain/content-specific parameters:** For example, limitations on text complexity for reading passages or length of conversation for interactive listening/speaking prompts.
2. **Administration conditions:** For example, for an online assessment, when or whether a student interacts directly with the assessment or when or whether the teacher administers the assessment directly to the student. This guidance may depend on test design decisions applied to each measurement target. For example, if the target focused on pre-symbolic communication, or focused on symbolic communication but would introduce too much abstraction in an on-screen, two-dimensional environment, the template might specify administration using concrete materials.
3. **Item format(s):** Some assessments may have a single item format for the domain; for example, all listening items are three-choice multiple choice items. Others may have a mix of item formats depending upon the measurement target, and this feature would be specified at the item template level.
4. **Measurement target(s):** The measurement target specifies what part of the claim (or standard) is being assessed. For example, if the writing standard is "support opinions, thoughts, ideas, and topics with details/examples", the measurement target may be "include at least one fact to support an opinion." For constructed-response items, the measurement target is usually included in and guides the scoring rubric.
5. **Limitations/parameters around measurement target:** For example, for the writing standard "support opinions, thoughts, ideas, and topics with details/examples," the opinion requested must be related to a familiar environment, for example, classroom/school activities.
6. **Evidence descriptions:** For example, a holistic scoring rubric for full credit may include "the student includes at least one reason the student holds this opinion;" for partial credit, the rubric may include "the student states an opinion but does not provide a reason for the opinion."
7. **Stimulus specifications/limitations:** These may be specified at the assessment, domain, or item template level and are related to feature 5. For example, tables must have no more than three rows and four columns; science passages should be accompanied by pictures and be no more than 50 words long; scenarios must be of common activities likely to be familiar to the student.
8. **Stem specifications/limitations:** For example, for a specific speaking item template, the stem must be in the form of a direct question; for a specific listening item template, the stem must clearly describe the expected interaction/response.
9. **Response specifications/limitations:** These may relate to form or content. For example, the distractors in a multiple choice item should draw on specified common misconceptions; or in an item that has the student select one or more pictures, once a student selects a picture, its frame should be highlighted.

(Figure continues on next page.)

10 **Item complexity/complexities/levels:** Depending on the structure of the standards, a single item template may address a single level or multiple levels of achievement, or allow for multiple levels of complexity. For example, a constructed-response writing task with a 3-point scoring rubric might address more than one level of achievement and, depending on the rubric, more than one level of complexity. A single template may be used to model items at several levels of complexity.

11 **Scoring rules/rubrics:** At times, the scoring rule is evident; for example, for single-correct multiple choice items, correct answers are usually worth 1 point. Other times, the nature of the standard and associated item will drive the scoring rule. For example, the student must correctly select two options to receive 1 point; the student response is scored according to the rubric in Figure 2 common to all items developed from the template.

12 **Ancillary and precursor skills required:** Ancillary are construct-irrelevant skills and often related to the administration format. Identifying these skills in the "standard" item can inform the development of accessibility tools and alternate modes of administration and response. These ancillary skills may be specified at the domain or test level. For example, if a needed ancillary skill is the ability to use a standard keyboard, options for administration may include using a scribe, an adaptive keyboard, or a communication device. A precursor skill is construct-relevant but hypothesized to come before the target in a typical learning sequence. For example, a precursor skill may be the ability to write a multi-sentence response if the standard requires supporting an opinion with at least one detail.

13 **Universal design for learning specifications and accessibility considerations:** These may be written at the test or domain level and augmented in the item template with more specifics if necessary. For example, the reading domain may specify that student-controlled read-aloud or automatic read-aloud must be available, and that the student (or administrator) can control the speed of the read-aloud; the speaking domain may specify that students can use assistive devices used in the classroom to produce speech; the listening domain may specify that the student's usual interpreter can sign the items and directions to the student using the student's usual mode for receiving oral communications; the writing domain may specify that students can use assistive devices or a scribe to produce text. These domain-level considerations will affect item templates differently, depending on other factors related to the items. At the item template level, specific instructions such as which manipulatives can be used and how they can be used should be included as needed.

14 **Supports allowable/available**: These may be specified for the test, domain, modality, and/or item. For example, specific accommodations are typically written for the test or domain as a whole. Supports such as the administrator reminding students about the nature of the task demand—for example, "A main idea is what the story is about"—may be at the item template level.

15 **Alternate modes or substitute items:** As discussed above, each item template should include information about how the item will be made available to students with various accessibility needs, if this attribute is not already covered by domain or test level features. Examples include having all options read aloud, or allowing a scribe for text-based responses. At the item template level, any need for substitute items and the form of those substitutes should be specified. For example, if a picture-based item cannot be delivered in braille, template features for a substitute item should be specified. Another example of an alternate mode would be the use of signing for items on the listening test.

A central requirement for test design, including item template development, is the involvement of people with the right expertise at various stages, including educators and others who will use and interpret the results of the test. In particular, educators who are familiar with the student population should be active in and advise at all stages of the evidence-centered design process. Any overall considerations for template development such as text complexity, text presentation, and administration conditions should be designed with the advice of measurement and content experts as well educators familiar with the population. Criteria for good item templates and draft item templates should be reviewed by educators with expertise in measurement, in English language development, and in education of the targeted student population. Those same educators should revise the criteria revised as needed. When policymakers are brought into the design process, either for the purpose of informing them or for collecting input or approval, educators with expertise in educating English learners with significant cognitive disabilities should be included, if possible.

Sample Item Templates

The templates below are based on one of the sample priority key concepts from the ALTELLA standards prioritization activity for reading, "Identify a main topic from a variety of texts." The first template and its assumptions are annotated with the numbers associated with the features listed above (Figure 1).

Assumptions

The templates are illustrations based on several assumptions about overall test design:

A. The priority key concept is equivalent to a standard, and each standard has three associated measurement targets at varying levels of complexity.[2] The test is designed for items to be written separately, not as a multi-item performance task with items administered in decreasing complexity or with increasing supports.

B. Complexity varies according to the cognitive performance expected of the student and according to the length of the text.

C. Test development guidelines already include general rules limiting the complexity of text (e.g., no compound sentences, avoiding pronouns).

1

[2] Given the multiple audiences for this report, for illustrative purposes the complexity is expressed in a four-point taxonomy (Webb, Alt, Ely, Cormier, & Vesperman, 2005). Other taxonomies may be more appropriate for alternate assessments.

Stimulus specifications in the template cover other constraints on complexity (i.e., text length and overall reading level).[3]

D. Documented separately, test specifications indicate whether a passage should support multiple items that measure different standards. If multiple items are to be written for a single passage, the template provides a cross-reference to the other standard(s). 1

E. Test designers provide a separate list of assumptions about how students may respond and what teachers may do during test administration. For example, students may indicate responses using any means of expressive communication (i.e., direct selection on screen, verbal response, gestures, sign, via an augmentative and alternative communication device, eye gaze). When students cannot indicate selections directly in the online platform, teachers may enter responses as the students indicate to them. The teacher may prompt the student to attend to the on-screen content but may not use any other prompts. When more than one item is associated with a passage, the text may be read in its entirety for each question. 2 13

F. There is a common list of accessibility features that are general to the reading domain. This list is separate, so that the accessibility section of the template is used to note any exceptions to the items on the general list, which includes:

    a. The student may use synthetic or human read-aloud, or have the text signed, for the passage, questions, and answer choices.

    b. The student may use all display features built into the user interface, including magnification, reverse contrast, color contrast.

    c. The student may use any familiar special equipment needed to interact with the computer (switches, adaptive keyboard/mouse).

    d. For offline items (those administered directly by the teacher), the teacher may use partner-assisted scanning (i.e., providing students each choice one at a time so the student can use a preferred communication mode to indicate responses to each choice).

With these general design assumptions in mind, Figure 2 provides a sample template for a multiple-choice item. An item template will usually include one or more sample items or point the item developer to sample items based on the template.

[3] Given this report's multiple audiences, for illustrative purposes the reading level is expressed in terms of lexiles. This measurement may not be the most appropriate for alternate assessments.

Figure 2: Sample template for grade 5 reading standard, multiple choice item

Note: Annotation numbers refer to the list of potential features of item templates in Figure 1.

Domain: Reading	**Standard: R.5-3: Identify main topics across content areas**
Target: Identify main idea in an informational text 4	
Target constraints: Main idea is explicitly stated in the text 5	
Accessibility: All accessibility features for reading domain apply 13	
Level 3: Identify main idea in an informational text 10	
Evidence: student selects the main idea 6	
Format: single-select multiple choice 3	Administration: online 2
Complexity: DOK level 1 10	Scoring: 0 (incorrect) or 1 (correct) 11
Stimulus constraints: Passage on an informational topic related to science or social studies; maximum length 40 words; <BR100 lexile. The passage should be three or four sentences in length. 7	
Stem constraints: Phrased as a question, uses term "main idea." 8	
Response constraints: Student selects the main idea from an array of three choices. Each choice is a sentence from the passage. 9	
Allowable supports: Teacher may define specific words for the student upon request. 14	
Alternate modes/items: braille 15	

(Figure continues on next page.)

Level 2: Identify the topic of an informational text	
Evidence: student selects the topic	
Format: single-select multiple choice	Administration: online
Complexity: DOK level 1	Scoring: 0 (incorrect) or 1 (correct)
Stimulus constraints: Passage on an informational topic related to science or social studies; maximum length 40 words; <BR100 lexile. The passage should be 2-3 sentences in length.	
Stem constraints: Phrased as a question: “What is this story about?”	
Response constraints: Student selects the topic from an array of three choices.	
Allowable supports: Teacher may define specific words for the student upon request.	
Alternate modes/items: braille	
Level 1: Identify object associated with an informational text	
Evidence: student identifies an object	
Format: single-select multiple choice	Administration: offline
Complexity: DOK level 1	Scoring: 0 (incorrect) or 1 (correct)
Stimulus constraints: Passage on an informational topic related to science or social studies and about a topic for which objects would commonly be available in a school. Maximum length 40 words; <BR100 lexile. The passage should be two or three sentences in length.	
Stem constraints: Phrased as a statement.	
Response constraints: Student selects the object from an array of three choices. The order in which objects are presented must be stated in the teacher instructions.	
Allowable supports: Teacher may read the text a second time.	
Alternate modes/items: N/A	

Figure 3 is an excerpt from a template based on the same standard, but with the expectation that the item is a technology-enhanced item type that allows the student to select text within the passage. The online platform is designed to allow selection at the sentence or word level. Parts of the template that vary from the multiple choice example in Figure 3 are underlined.

Figure 3: Excerpt from sample template for grade 5 reading standard technology-enhanced item	
Domain: Reading	**Standard: R.5-3: Identify main topics across content areas**
Target: Identify main idea in an informational text	
Target constraints: Main idea is explicitly stated in the text.	
Accessibility: All accessibility features for reading domain apply.	
Level 3: Identify main idea in an informational text	
Evidence: student selects the main idea	
Format: single-select multiple choice	Administration: online
Complexity: DOK level 1	Scoring: 0 (incorrect) or 1 (correct)
Format: select text	Administration: online
Stimulus constraints: Passage on an informational topic related to science or social studies; maximum length three sentences with the main idea expressed in just one sentence. <BR100 lexile	
Stem constraints: Stem wording is “Choose the main idea.”	
Response constraints: Student selects the sentence that conveys the main idea.	

Conclusion

The field of alternate English language proficiency assessment is relatively new and the considerations for item templates provided here are based on current knowledge. Research, such as cognitive labs, is needed to support the use of item templates that allow English learners with significant cognitive disabilities to show what they know and can do.

As the field learns more about English learners with significant cognitive disabilities and their interactions with assessment items, researchers and developers can built upon tis foundation for continuous improvement. Through ALTELLA and other future efforts, English learners with significant cognitive disabilities may be appropriately served to develop proficiency in English and to further their post-secondary success in college, community, and career.

References

Cameto, R., Haertel, G., DeBarger, A., & Morrison, K. (2010). *Applying evidence-centered Design to alternate assessments in mathematics for students with significant cognitive disabilities* (Alternate Assessment Design-Mathematics Technical Report #1: Project Overview). Menlo Park, CA: SRI International.

Center for Applied Special Technology. (2015). *Top 10 UDL tips for assessment*. Wakefield, MA: Author. Retrieved from http://castprofessionallearning.org/project/top-10-udl-tips-for-assessment

Center for Applied Special Technology. (2018). *Universal design for learning guidelines version 2.2* [graphic organizer]. Wakefield, MA: Author. Retrieved from http://udlguidelines.cast.org/more/downloads

Christensen, L. L., Gholson, M. L., & Shyyan, V. V. (2018, April). *Establishing a definition of English learners with significant cognitive disabilities* (ALTELLA Brief No. 1). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research, Alternate English Language Learning Assessment project: http://altella.wceruw.org/resources.html

Christensen, L. L., Mitchell, J. D., Shyyan, V. V., & Ryan, S. (2018, September). *Characteristics of English learners with significant cognitive disabilities.* Madison, WI: University of Wisconsin–Madison, Alternate English Language Learning Assessment (ALTELLA). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research: http://altella.wceruw.org/resources.html

Christensen, L. L. & Mitchell, J. D. (2018, September). *Classroom perspectives on English learners with significant cognitive disabilities*. Madison, WI: University of Wisconsin–Madison, Alternate English Language Learning Assessment (ALTELLA). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research: http://altella.wceruw.org/resources.html

Christensen, L. L., Shyyan, V., & Johnstone, C. (2014). Universal design considerations for technology-based, large-scale, next-generation assessments. International Dyslexia Association: Perspectives on language and literacy, 23–31.

Clark A. K., Karvonen, M., Wells-Moreaux, S., & Erickson K. (2018, April). *Characteristics and performance of students with significant cognitive disabilities who may be English learners.* Paper presented at annual meeting of the American Educational Research Association, New York, NY. Retrieved from https://dynamiclearningmaps.org/sites/default/files/documents/presentations/AERA_2018_EL_Students_with_SCD.pdf

Gordon, D. T., Gravel, J. W., & Schifter, L. A. (2010, April). *Perspectives on UDL and assessment: An interview with Robert Mislevy*. Wakefield, MA: National Center on UDL. Retrieved from http://www.udlcenter.org/sites/udlcenter.org/files/mislevy_final_0.pdf

Hauck, M. C., Pooler, E., & Anderson, D. P. (2015, May). *ELPA21 item development process report*. Washington, DC: Council of Chief State School Officers.

Johnstone, C. J., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.umn.edu/docs/OnlinePubs/StateGuideUD/UDmanual.pdf

Johnstone, C., Altman, J., Thurlow, M., & Moore, M. (2006). *Universal design online manual*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.info/Resources/publications/UDmanual/default.html

Mislevy, R. J. (2011). *Evidence-centered design for simulation-based assessment* (CRESST Report 800). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Mislevy, R. J., & Haertel, G. (2006). Implications for evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice, 25*, 6–20.

Riconscente, M. M., Mislevy, R. J., & Hamel, L. (2005). *An introduction to PADI task templates* (Principled Assessment Designs for Inquiry Technical Report 3). Menlo Park, CA: SRI International.

Rooney, P. (2017, June 28). *Update on English language proficiency assessments. Memo to state assessment directors and state Title I directors*. Washington DC: U.S. Department of Education. Retrieved from https://www2.ed.gov/admins/lead/account/saa/elpassessments72117.pdf

Thurlow, M., Christensen, L., & Shyyan, V. (2016). *White report on English language learners with significant cognitive disabilities*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, English Language Proficiency Assessment for the 21st Century.

Webb, N. L., Alt M., Ely, R., Cormier, M., & Vesperman, B. (2005, December). The web alignment tool: development, refinement, and dissemination. In *Aligning assessment to guide the learning of all students: 6 reports*. Washington, DC: Council of Chief State School Officers.

ALTERNATE ENGLISH LANGUAGE
LEARNING ASSESSMENT PROJECT

Wisconsin Center for Education Research | University of Wisconsin–Madison
1025 West Johnson Street | Madison, WI 53706 | altella.wceruw.org