



ALTERNATE ENGLISH LANGUAGE
LEARNING ASSESSMENT PROJECT



Developing a Request for Proposals for an Alternate Assessment of English Language Proficiency

ALTELLA Report

September 2018

Laurene L. Christensen, Vitaliy V. Shyyan, and Kristopher Stewart



Wisconsin Center for
Education Research
UNIVERSITY OF WISCONSIN-MADISON

Developing a Request for Proposals for an Alternate Assessment of English Language Proficiency

ALTELLA Report

September 2018

Laurene L. Christensen, Vitaliy V. Shyyan, and Kristopher Stewart

Suggested Citation

Christensen, L. L., Shyyan, V. V., & Stewart, K. (2018, September). *Developing a request for proposals for an alternate assessment 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>

© 2018 by Laurene L. Christensen, Vitaliy V. Shyyan, and Kristopher Stewart. All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as suggested. The contents of this report were developed under a grant from the U.S. Department of Education. However, those contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal government.

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.

The ALTELLA project is a partnership of five state departments of education and the Wisconsin Center for Education Research at the University of Wisconsin–Madison. This collaboration 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	Melissa Gholson	Sarah Ryan
Kristen Burton	Jonathan Gibson	Kristopher Stewart
Indira Ceylan	Miguel Hernandez	Sonia Upton
Elizabeth Cranley	James Mitchell	Cha Kai Yang

Publication Coordinator Indira Ceylan

Report Editor Karen Faster

Graphic Designer Janet Trembley

ALTELLA Request for Proposals Activity Participants

Marlene Johnston, Arizona Department of Education
Bethany Zimmerman, Arizona Department of Education
John Jacquith, Michigan Department of Education
Jennifer Paul, Michigan Department of Education
Tracy Montez Lindner, Minnesota Department of Education
Nicole Adams, South Carolina Department of Education
Mami Itamochi, West Virginia Department of Education
Sonja Phillips, West Virginia Department of Education

Wesley Bruce, Independent Consultant

Margaret Ho, English Language Proficiency Assessment for the 21st Century (ELPA21)

Table of Contents

Introduction	1
Background	2
Methods	2
Results	4
Discussion	9
Conclusion.....	9
References	10

List of Tables

Table 1: Multi-attribute consensus building results: Item development.....	5
Table 2: Multi-attribute consensus building results: Field test	8

Introduction

On July 18, 2014, the U.S. Department of Education released guidance relating to English learners with disabilities. Included in this guidance was the clarification that all English learners, including those with the most significant cognitive disabilities, must participate in a state's English language proficiency assessment. If a student is unable to participate in the general content assessment with accommodations, then the state must provide an alternate assessment of English language proficiency.

In the context of increasing attention to improving services for English learners who have significant cognitive disabilities, the importance nationwide for providing alternate English language proficiency assessments increases as well. As of September 2018, members of the WIDA Consortium (i.e., an assessment consortium of 39 states and U.S. territories) administer the Alternate ACCESS for ELLs to English learners with significant cognitive disabilities. In the spring of 2018, the Texas Education Agency piloted an alternate assessment of English language proficiency. However, the remaining states and territories do not have an operational assessment of English language proficiency that is appropriate for English learners with significant cognitive disabilities. ELPA21 (i.e., an assessment consortium of eight states) is in the design phase of their alternate English language proficiency assessment.

Putting together a request for proposals is a critical step in the development of a new assessment. Martineau and Domaleski (2018) note that a "good [request for proposals] is a complex document, reflecting the complexity of the projects it is intended to help states procure" (p. 1). Requests for proposals provide an opportunity for a state or an organization to describe what they value and expect from the assessment. Requests for proposals are also a key arena to clearly define roles, provide a statement of work, and outline the terms and conditions for the work. Crafting an alternate assessment of English language proficiency brings together the important elements of both English language proficiency assessment and alternate assessment. In planning for the development of such an assessment, collaboration among experts in the fields of English language development, special education, and alternate assessment is critical so that the values and expectations inherent in both English language proficiency assessment and alternate assessment are included. It may also be beneficial for states and territories to work together on such an assessment. The overall population of students is small, and by working together, all stakeholders can learn more about effective assessment approaches that provide valid and reliable information about the language development of English learners with significant cognitive disabilities.

The Alternate English Language Learning Assessment (ALTELLA) project researches instructional practices, multi-tiered accessibility features and accommodations, and assessment approaches to use with English learners with significant cognitive disabilities. ALTELLA supports an evidence-centered design approach that informs the project's understanding of alternate English language proficiency assessment for these students. This report provides considerations for organizations that seek to develop requests for proposals for developing alternate English language proficiency assessments, with the focus on item development and field testing processes and procedures.

Background

The Center for Assessment has developed a request for proposals toolkit that includes a user guide and a model outline that can be adapted to a variety of assessment needs (Martineau & Domaleski, 2018). The goal of the request for proposals activity for the ALTELLA project was to develop a list of customized considerations for states and territories as they embark on creating a request for proposals for an alternate English language proficiency assessment. The diverse nature of the population of English learners with significant cognitive disabilities means a request for proposals must consider these students' needs so that the resulting assessment is a valid and reliable measure that supports instructional services for these students in English. ALTELLA's methodology to identify and prioritize potential elements of a request for proposals and the findings described here may be useful to those stakeholders who are developing a request for proposals related to an alternate assessment of English language proficiency.

Methods

To develop a list of prioritized considerations that support generating a request for proposals addressing two core aspects (item development and field testing), ALTELLA engaged participants in a multi-attribute consensus building process. This research method is a quantitative approach for determining a group's opinion about the importance of each variable on a list (Shyyan, 2018; Shyyan, Christensen, Thurlow, & Lazarus, 2013; Vanderwood, Ysseldyke, & Thurlow, 1993). This method enables a small or large group of participants to generate and discuss a set of items, weight the importance of each item, and debrief their weightings to reach consensus or identify and document the sources of differences in participants' perceptions. According to Lewis and Johnson (2000), multi-attribute consensus building is appropriate when judgments are required as part of the decision-making process, and it allows for the decision process for an individual or a group of individuals who consider several alternatives. This research method is grounded in participants' considerations of a set of variables, in this case, related to the process of developing a successful request for proposals.

For ALTELLA's multi-attribute consensus building process, participants were asked to discuss and weight various elements to be included in a request for proposals for item development and field testing components of an alternate assessment of English language proficiency. The multi-attribute consensus building method allowed participants to consider a list of core elements of sample request for proposals stemming from the Center for Assessment guidelines (Martineau & Domaleski, 2018), generate additional elements, and weight their importance on a scale from 1–20 (very unimportant), to 21–40 (unimportant), to 41–60 (neither unimportant nor important), to 61–80 (important), to 81–100 (very important), provided that *at least* one element was assigned the maximum 100 weighting. After all participants finished discussing and generating their weightings, each reported her or his weightings for each request for proposals topic. All weightings were instantly entered into a multi-attribute consensus building spreadsheet and projected onto a screen, and, if preferred, participants were able to change their own weighting at any point of the discussion to achieve a higher level of consensus. The spreadsheet (which can be downloaded at <https://nceo.umn.edu/docs/Tools/MACBspreadsheetBlank.xlsx>) automatically calculated ranges, overall proportional weights, and importance averages for each element of a request for proposals. The whole group subsequently discussed these findings, especially the elements that received more varied weightings. Participants could change any of their weightings at during the general discussion if they became convinced by others of a different level of importance of a given element of a request for proposals.

The multi-attribute consensus building process was conducted once for the topic of item development and replicated for the topic of field testing. This process was conducted in two sessions. The first session, focusing on item development, was completed as part of an in-person meeting held in Phoenix, Arizona, on April 24, 2018. The group started exploring field testing, then reconvened during a virtual meeting on July 24, 2018 to complete the multi-attribute consensus building process.

Participants included representatives from the five partner states (Arizona, Michigan, Minnesota, South Carolina, and West Virginia). Two additional project experts participated in the multi-attribute consensus building activity on the item development topic. Participants had background in special education, English language development, and assessment development.

Participants received a list of general considerations, or elements, related to both item development and field testing; this list was drawn from the information in the Center for Assessment request for proposals template (Martineau & Domaleski, 2018). Participants then discussed the elements on the list and were given the opportunity to refine and add items.

Results

In general, nearly all of the elements in both consensus building activities were rated as very important by the participants. In both activities, even the lowest rated elements were rated within the range of important. States and organizations considering a request for proposals related to the development of items and field testing an alternate assessment of English language proficiency, may find useful thinking through all the list considerations for potential inclusion in a request for proposals.

The highest rated elements in the item development multi-attribute consensus building (Table 1) were:

- Items have to be aligned to alternate ELP standard.
- Items are amenable to accessibility features and accommodations, including assistive technology.
- The criteria in the item templates will be adhered to.
- Universal design for learning principles are adhered to during item development.

The lowest rated element in the item development multi-attribute consensus building was “Conduct an independent study of item alignment”; however, this element was still rated as important by the group. Participants noted that this element seemed less important than other elements in the set; they noted that this consideration may be costly.

Table 1: Multi-attribute consensus building results: Item development

Item Development Topic	Range of Ratings	Proportional Weight	Average of Weightings
Service providers offer trainings on the item writing process.	84-100	0.16	90
Item writers include accessibility, bias, and sensitivity experts.	85-100	0.17	93
Items undergo a quality control process, including an editorial review.	75-100	0.16	90
Items are written in collaboration with educators who have firsthand experience with content and English learners with significant cognitive disabilities.	60-100	0.16	89
Items are amenable to accessibility features and accommodations, including assistive technology.	90-100	0.17	96
Appropriately designed cognitive labs are conducted with English learners with significant cognitive disabilities with a range of language and disability needs to gather information on item templates.	80-98	0.17	91
Items are reviewed by educators who have firsthand experience with content and instruction of English learners with significant cognitive disabilities.	85-100	0.17	94
Items have to be aligned to alternate English language proficiency standards.	90-100	0.18	98
Items comply with accessible portable item protocol (APIP)/question and test interoperability (QTI) standards.	80-100	0.17	93
The criteria in the item templates will be adhered to	90-100	0.18	96
Universal design for learning principles are adhered to during item development.	90-100	0.17	96
Professionally developed media resources are provided.	80-100	0.17	91
Determine totals of items, proportion of items, item types, item specifications, difficulty ranges for the development process and subsequent gap analyses.	82-100	0.17	93
Have a range items responsive to language- and disability-related needs to be considered at entry points (personal needs profile/individual student assessment accessibility profile-like approach).	79-100	0.17	94

(Table continues on next page.)

Item Development Topic	Range of Ratings	Proportional Weight	Average of Weightings
Have a range items responsive to language- and disability-related needs to be considered at entry points (personal needs profile/individual student assessment accessibility profile-like approach).	79-100	0.17	94
Conduct an independent study of item alignment.	15-100	0.12	67
Have a comprehensive, high-quality metadata system to support item development (including state-specific needs, link to alternate English language proficiency assessment standards).	75-100	0.17	92
Items should be designed considering all modes of administration (computer-based, paper-pencil, braille forms, etc.).	70-100	0.17	93
Have a rigorous quality control process in place to reduce the numbers of rejected items and increase the numbers of accepted/usable items.	67-100	0.16	86
Ensure that the length of generated/editable stimuli and items is aligned to different complexity levels (linguistic complexity); authentic content needs to allow permissions to edit.	80-99	0.17	92
Items are owned by the entity that releases the request for proposals.	78-100	0.17	91
Items undergo accessibility, bias (avoiding regionalisms, gender, other cultural dimensions), sensitivity (topics to be avoided and treated with caution), and linguistic complexity reviews.	70-100	0.16	90
An ongoing item development process is in place.	45-100	0.15	84
Item writing and review meetings and products are Americans with Disabilities Act (ADA)/Section 508 compliant.	79-100	0.16	89
Conduct usability studies with different types of students (item functionality).	70-100	0.16	86

The highest rated elements in the field testing multi-attribute consensus building (Table 2) were:

- The field test includes all item types.
- Train field test administrators and other educators involved (platform considerations).
- Comparability studies across different modes of presentation.

The lowest rated element in the field test multi-attribute consensus building was “Field test above grade level.” In relation to this element, some participants inquired whether testing should be at the student’s grade or developmental level; the importance of testing at the student’s grade level was emphasized. Participants also noted that testing above grade level might be needed for developing a vertical scale.

Table 2: Multi-attribute consensus building results: Field test

Field Test Topic	Range of Ratings	Proportional Weight	Average of Weightings
The field test includes all item types.	100-100	0.17	100
Develop a sampling plan to address students' range of characteristics and needs for the field test.	90-100	0.17	95
Conditions are identified under which embedded field testing is not appropriate; all effort should be made to embed in the next rounds of field-testing.	90-98	0.16	92
The field test includes all accessibility features and accommodations.	90-100	0.17	97
Content and data reviews are conducted before and after the field test.	90-100	0.16	94
Include non-English learner students with significant cognitive disabilities in the field test.	90-100	0.17	96
Develop a plan for field-testing and re-field-testing.	90-100	0.17	96
Identify populations that may be challenging to field test with due to cost implications, low incidence.	90-93	0.16	91
Field test at grade level.	90-100	0.17	95
The test administration manual needs to fully document the field test process.	89-100	0.16	92
Field test above grade level.	0-85	0.07	41
Develop a comprehensive field test schedule.	90-100	0.17	95
Train field test administrators and other educators involved (platform considerations).	90-100	0.17	98
Develop a practice test.	90-100	0.17	96
Develop the vertical scale during the first year.	90-100	0.16	93
Platform needs to be compatible with assistive technology devices and other supports.	90-100	0.17	96
Comparability studies across different modes of presentation.	95-100	0.17	98

Discussion

This multi-attribute consensus building process was designed to generate recommendations from the states participating in ALTELLA about what elements of a request for proposal are important and need to be considered in this process. The participants in this activity represent three assessment systems (AZELLA, ELPA21, and WIDA) focusing on the considerations for a request for proposals that they perceived to be critical for collaborating with service providers on item development and field test approaches. Although the lists of elements are not exhaustive, and there are additional components to a request for proposals beyond item development and field testing, participants perceived the lists to be central in structuring these activities. This agreement suggests that nearly all of the elements of a request for proposals are critical and should not be excluded from such a document. However, this finding may not be a surprising finding given the overall importance of the request for proposals in conveying the level of specificity in what should be included in a plan for developing a new assessment.

Conclusion

The field of alternate assessment of English language proficiency continues to grow, and educators, researchers, and policy makers are establishing the best practices in understanding how English learners with significant cognitive disabilities demonstrate their growth in English language development. Among the available resources for test development is the Assessment RFP Development Toolkit – Model RFP Outline developed by Martineau (2018). This outline should be customized based on critical elements of alternate assessment and English language proficiency assessment such as those included in this report. As understanding continues to grow, the critical components of the assessment design and implementation that are necessary in a request for proposals for this specific type of assessment will be uncovered and further refined.

References

- Lewis, D. R., & Johnson, D. R. (2000). *Participatory evaluation for special education and rehabilitation*. Washington DC: American Association on Mental Retardation.
- Martineau, J. (2018). Assessment RFP development toolkit – Model RFP outline. Dover, NH: Center for Assessment. Retrieved from <https://www.nciea.org/sites/default/files/inline-files/RFP%20Toolkit%20-%20Model%20RFP%20Outline.docx>
- Martineau, J., & Domaleski, C. (2018). Development toolkit for procuring assessment services. Dover, NH: Center for Assessment. Retrieved from <https://www.nciea.org/sites/default/files/inline-files/RFP%20Toolkit%20-%20User%20Guide.pdf>
- Shyyan, V. (2018). Instructional strategies for developing intercultural competence in the language classroom. *Notos*, 14(1).
- Shyyan, V., Christensen, L., Thurlow, M., & Lazarus, S. (2013). Multi-attribute consensus building tool. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Vanderwood, M. L., & Erickson, R. (1994). Consensus building. *Special Services in the Schools*. 9(2), 99–113.
- Vanderwood, M., Ysseldyke, J., & Thurlow, M. L. (1993). *Consensus building: A process for selecting educational outcomes and educators*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



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