

Research Notes



English Language Proficiency Standards and Assessment Alignment

Issues, Insights, and Innovations
for Guiding Peer Review

Alison L. Bailey, Mikyung Kim Wolf, and Laura Ballard



This research note is based on our experience with language complexity protocol development and our research findings from the William T. Grant Foundation-funded project, *An Investigation of Language Demands in Standards, Assessments, and Curricular Materials for English Learners* (full report forthcoming). A major goal of this project is to assist in the reduction of inequality in educational experiences that English learner (EL) students experience in U.S. schools by investigating the language demands embedded in academic content standards and English language proficiency (ELP) standards, their linked assessments, and curricular materials in use with EL students in the fifth grade.¹ In doing so, we aim to identify essential language knowledge and skills that EL students need to acquire to meaningfully participate and succeed in academic learning and assessment. This work comes at a critical juncture in the field of educational assessment; the hiatus and adjustments to student testing during the coronavirus pandemic have raised questions about testing's role in instruction. This is coupled with serious concerns about fairness and the consequences of testing for culturally and linguistically diverse learners in the wake of the racial justice reckoning experienced across all sectors of society. Ensuring that assessments are appropriately developed and validly used for cultivating equitable education systems is crucial, now more than ever.

The research note focuses on the alignment aspect of ELP assessments, one of the required types of validity evidence for the federal peer review process of states' assessment systems. A basic tenant of current U.S. education policy is the alignment between what a test assesses and what content has been determined as meaningful for students in a given grade. A student's performance on a well-aligned assessment should indicate that student's level of mastery of content

knowledge. Whereas the alignment methods and literature on academic content assessments are relatively well-established, little guidance is available as to how to evaluate the alignment of ELP assessments to standards in terms of their content match (i.e., overlap in language demands between assessments and standards).

During the project investigation, we have developed a principled way to characterize language complexity inherent in standards, assessment, and curricular materials. We use the term "language complexity" to be inclusive of the complexity of linguistic forms and processing skills that contribute to language demands imposed on students. This research note particularly deals with high-level methodological guidance in two areas: First, we make suggestions for operationalizing the constructs of language complexity and cognitive complexity in this context. Second, we make recommendations for questions states might consider as they or their contractors assemble evidence of alignment pertaining to ELP standards and ELP assessments, as well as correspondence between ELP standards and academic content standards.

Intended audience and purpose

The intended audience of this research note spans state education agencies (SEAs), assessment consortia, and assessment developers who must provide evidence of alignment for the federal government peer review processes (U.S. Department of Education [USDOE], 2018) outlined in the following pages. A secondary audience is third-party vendors and consultants contracted to perform alignment studies. Additionally, agencies of the federal government and their selected peer reviewers may find the research note useful in their upcoming reviews and in future directions to states.

This research note has three objectives:

1. Review the statutory requirements of the USDOE's assessment peer review process, particularly regarding alignment of the state's ELP assessment to its standards (see Box 1) and describe the significance of alignment evaluation in educational policy and for effectiveness of EL student education in particular.
2. Share insights from research that add to our understanding of the language demands in standards and assessments based on the development of the language complexity analysis protocol.
3. Outline key considerations for evaluating alignment between ELP standards and ELP assessments and for establishing correspondence between ELP standards and academic content standards, namely for English language arts and mathematics.

Objective 1: Why this matters for the assessment peer review process

The principled approach we have taken to characterizing language demands in assessments, standards, and curricular materials has raised practical issues and has allowed for various insights and innovations that may prove useful for states and state contractors in conducting correspondence and alignment studies. A state's annual ELP assessment is designed to measure an EL students' proficiency in the English language. In the reauthorized Elementary and Secondary Education Act, the Every Student Succeeds Act (ESSA, 2015), "ELP assessments must be aligned to the ELP standards (Section 1111(b)(2)(G)) and measure ELs' proficiency levels annually in the four recognized

Box 1. Assessment Peer Review Process: Section 3: Technical Quality—Validity Critical Element 3.1—Overall Validity, Including Validity Based on Content

Examples of Evidence

Evidence to document adequate validity based on content for the State's general assessments includes:

- Validity evidence based on the assessment content that shows levels of validity generally consistent with expectations of current professional standards, such as:

[...]

- For the ELP assessments, *expert review of items showing that the items address language demands of grade level academic content standards.*
- Evidence of *alignment*, including:
 - Report of results of an independent alignment study that is technically sound (i.e., method and process, appropriate units of analysis, clear criteria) and documents adequate alignment, specifically that:
 - *Each assessment is aligned to its test blueprint, and each blueprint addresses: (1) depth and breadth of the State's academic content standards; or (2) the depth and breadth of the State's ELP standards;*
 - *;* [sic]
 - *The State follows procedures to ensure alignment during test development;*
 - Description of a systematic process and timeline the State will implement to address any gaps or weaknesses identified in the alignment studies. [Emphasis in italics added].

From *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process* by U.S. Department of Education, 2018, pp. 47–48 (<https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>). In the public domain.

domains of language skills: speaking, listening, reading, and writing (Section 1111(b)(1)(F))” (USDOE, 2018, p. 25). Moreover, those ELP standards must correspond to a state’s academic content standards in English language arts, mathematics, and science (USDOE, 2018, p. 30). Statutory requirements of the ESSA mean that states will need to document these alignment efforts as part of Section 3: Technical Quality—Validity (USDOE, 2018). *A State’s Guide to the U.S. Department of Education’s Assessment Peer Review Process (Assessment Peer Review Process, henceforth)* explicitly requires states to submit evidence of alignment as Critical Element 3.1 of validity based on assessment content as follows:

The State’s ELP assessments measure the knowledge and skills specified in the State’s ELP standards, including:

- Documentation of adequate alignment between the State’s ELP assessment and the ELP standards the assessment is designed to measure in terms of language knowledge and skills, the depth and breadth of the State’s ELP standards, across all proficiency levels, domains, and modalities identified therein;
- Documentation of alignment (as defined) between the State’s ELP standards and the language demands implied by, or explicitly stated in, the State’s academic content standards. (Critical Element 3.1—Overall Validity, Including Validity Based on Content, USDOE, 2018, p. 48).

Typically, alignment evaluation for content assessments entails the content match in the knowledge, skills, and abilities (KSAs) of academic content standards as well as the KSA coverage (i.e., breadth) and its cognitive complexity level (i.e., depth). In the case of ELP assessments, the aspects of alignment are more complex because the content of language KSAs are intertwined with cognitive complexity (Paradis et al.,

Box 2. Assessment Peer Review Process: Section 1: Statewide System of Standards and Assessments. Critical Element 1.2—Challenging Academic Content Standards/Coherent and Progressive ELP Standards that Correspond to the State’s Academic Content Standards

Examples of Evidence

- Evidence that the State’s ELP standards are appropriate and correspond to the State’s academic content standards includes:
 - Documentation that the *four language domains (speaking, listening, reading, and writing), separately and/or in an integrated fashion, are included in the standards.*
 - Demonstration of a strong correspondence or linkage between the State’s academic content standards and the State’s ELP standards, such that the State can claim that language requirements outlined in the ELP standards correspond with the academic language demands of the State’s academic content standards. This evidence does not need to demonstrate that ELP standards include knowledge, skills, or vocabulary from the State’s academic content standards.
 - A detailed description of the strategies the State used to ensure that its ELP standards adequately specify *English language knowledge and skills necessary to reflect the language needed to acquire and demonstrate the skills identified in the State’s academic content standards in at least reading/ language arts, mathematics, and science.*
 - Reports of external independent reviews of the State’s ELP standards, summaries of reviews by educators in the State, or other documentation. This documentation should confirm that the *State’s ELP standards represent the English language proficiency expectations needed for ELs to demonstrate their achievement of skills identified in the State’s academic content standards appropriate to each grade-level/grade-band in at least reading/ language arts, mathematics, and science.* [Emphasis in italics added].

From *A State’s Guide to the U.S. Department of Education’s Assessment Peer Review Process* by U.S. Department of Education, 2018, p. 30–31 (<https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>). In the public domain.

2017) and can be difficult to separate. Unlike academic content standards, ELP standards commonly delineate different levels of proficiency. Examining the alignment between ELP standards and ELP assessments in terms of levels of proficiency is a new dimension to consider.

There is a fundamental demand for the clear articulation of the English language knowledge and skills found in ELP standards so that these may be compared with the inherent language complexity of the academic content standards (whether overt or implicit) to ascertain the degree of correspondence between them and the ELP standards' degree of alignment with ELP assessment items. This is important because "ELP standards should contain language proficiency expectations that reflect the language needed for EL students to acquire and demonstrate their achievement of the knowledge and skills identified in the State's academic content standards appropriate to each grade in at least reading/language arts, mathematics, and science" (USDOE, 2018, p. 24).

There were efforts to make such characterizations and determine the correspondences and alignments between standards and assessments with a prior generation of ELP and academic content standards and assessments (e.g., Bailey et al., 2007; Chi et al., 2011). Now a new generation of ELP standards and assessments—such as the ELPA21 state collaborative ELP standards (Council of Chief State School Officers [CCSSO], 2014) and WIDA English language development standards framework (WIDA, 2014, 2020)²—require reestablishing the two components of content validity involving evidence demonstrating (a) alignment between ELP assessment and ELP standards (Critical Element 3.1) and (b) correspondence of language demands between ELP standards and academic content-area standards (ELP and academic content, Critical Element 1.2 and Critical Element 3.1; see Box 1 and Box 2 in this research note for acceptable

evidence of these two critical elements). However, there are likely to be states without in-house experience or expertise in conducting the linguistic analyses needed to demonstrate alignment and correspondence. Further, contractors to states have only infrequently been requested to document alignment of the language demands between ELP and academic content standards. One of the aims in the William T. Grant Foundation project was to provide a research-based perspective on language demands and inform new alignment and correspondence efforts intended for the improvement of EL student assessment.

Objective 2: Insights from research on language demand features in standards and assessments

Development of language demand protocols

In order to characterize language demands, including language complexity in standards and assessments, the project developed protocols encompassing a number of categories drawn from previous literature (see the appendix for a list of relevant literature). In developing and adopting various linguistic categories, a combination of systemic functional linguistics, sociocultural theory, and speech act theory comprises our main theoretical frameworks. In recognition of the multifaceted nature of language demands, a sample of categories included in the project protocols follows:

- Complexity of language forms (e.g., lexical, syntactic, and holistic linguistic complexity levels)
- Language functions (e.g., describing, explaining, comparing)
- Language use contexts (e.g., social/interpersonal, school-navigational, general-academic, and discipline-specific contexts)

- Genre of language tasks/types of practices (e.g., engaging in academic discussion, writing a research report)
- Complexity of language processing skills (e.g., foundational, basic, and higher order language skills)
- Discreteness versus integration of language modalities (e.g., listening vs. listening and speaking about a topic followed by writing about it)
- Pragmatic arrangements (e.g., communicative mode such as collaborative, interpretive, and productive; types of interaction; audience; speaker)

Some of the categories above were applied to both standard descriptions and assessment items while others were applied to assessment items only. For example, the complexity of language forms can be evaluated in a quantifiable way for concrete assessment materials (e.g., reading passages, listening stimuli, items). In contrast, standards tend to make general reference to grade-appropriate texts, making it impractical to evaluate specific linguistic complexity.

Among the categories listed, the pragmatic arrangements category, not emphasized in previous work, is worth noting. The pragmatic arrangements in our protocol are adapted from Bailey's (2015) initial articulation of the ways in which standards and other documents portray social-communicative interactions in school contexts. This use and organization of language (and nonlinguistic forms of communication) include the range of audiences students are expected to address (e.g., other students, teachers, counselors), how classroom participation is configured (e.g., one-to-one, one-to-many interactions), conversational turn-taking practices, and mention of linguistic and nonlinguistic resources that can be recognized as students' multicompetences (e.g., use of first language to access content, use of visual aids, graphics). The subcategories related to pragmatic arrangements

are intended to capture the paradigm focusing on communicative language use in current ELP standards. This approach has proven useful for capturing not only language knowledge and skills but also language use in authentic disciplinary practices that is also a salient feature of the current academic content standards. By having a set of pragmatic categories, alignment examination of the representation of language use in ELP assessments and ELP standards (and the academic content standards) can be systematically conducted.

Notably, the project did not create a singular rating of linguistic complexity by which to judge a match for language demands across different types of materials (i.e., content standards, ELP standards, ELP assessments, curricular materials). Rather, the project protocols operationalized language demands based on the wide array of features outlined above (see Figure 1 and Figure 2 for examples).

Application of language demand protocols

During the iterative process of development and application of the language demands protocols in our project, new insights emerged concerning (a) the intersection of language demand variables and cognitive complexity measures, (b) the unit of analysis, (c) the parsimony for coding categories, and (d) rater/coder qualifications. Herein we briefly describe our experience and insights for each point.

First, we often found a close intersection of a group of language demand variables with cognitive complexity measured using Webb's (2005) Depth of Knowledge scale. In particular, when analyzing the language demands in standards and assessments in terms of language functions and complexity of language processing skills, these categories are naturally interwoven with cognitive demands. For instance,

summarizing a reading passage with complex linguistic forms on an abstract topic will likely require high-level cognitive demand. Thus, language functions and complexity of language processing skills provide fine-grained level of language demands and help determine the cognitive demands of standards and assessments. Alternatively, the evaluation of cognitive complexity may be redundant when the detailed evaluation of language demand is performed.

Second, standards commonly describe multiple skills, including both a primary language feature as well as one or more secondary or subfeatures, which makes determining the unit of language analyses complicated. In our project, we undertook a detailed analysis that unpacked the standards language to list multiple embedded skills. Even if a standard was written to a specific language skill or modality (e.g.,

writing), the standard may implicitly require other modalities (e.g., reading). Thus, no standard was taken at face value but was analyzed for all possible demands on the learner. The exercise of standards descriptions and documentation of skills in each standard not only facilitated a common understanding of standards among coders but also clarified cross-modalities embedded in reading, writing, and speaking/listening standards.

Third, when analyzing the language demands in such a comprehensive way using the protocol categories listed previously, the evaluation of alignment may be conducted by a set of certain categories (e.g., the coverage and mapping of language functions, language skills, and integration of modalities). This type of fine-grained analysis of language demands requires technical training and a substantial time commitment.

FIGURE 1: Example coding of language demands within an English language proficiency (ELP) standard

Language Functions: identify, interpret, explain, summarize

Language Modality
Discrete: listening, reading,
Integrated: listening & reading

Pragmatic Arrangements
Communicative Mode: interpretive, productive
Speaker: undefined
Audience: undefined

Standard 4-5.1
An ELL can...

Construct meaning from **oral presentations** and literary and informational **text** through grade-appropriate **listening, reading, and viewing.**

Level 5

use a wide range of strategies to:

- **determine** two or more main ideas or themes
- **explain** how key details support the main ideas or themes
- **summarize** a **text from read-alouds, written texts, and oral presentations.**

Note: ELPA21 Standard 4-5.1, Level 5. The following coding categories are not applicable to standards and thus are not shown in the figure: Language form complexity, language use context, genre, and complexity of language processing skill. ELL = English language learner. Adapted from *English Language Proficiency (ELP) Standards With Correspondences to K-12 English Language Arts (ELA), Mathematics, and Science Practices, K-12 ELA Standards, and 6-12 Literacy Standards* by Council of Chief State School Officers (p. 18), 2014 (https://ccsso.org/sites/default/files/2017-11/Final%204_30%20ELPA21%20Standards%281%29.pdf). Content copyright 2014 by the Council of Chief State School Officers.

FIGURE 2: Example coding of language demands within an English language proficiency (ELP) assessment task

Language Form Complexity Lexical: low Syntactic: moderate Holistic: moderate	
Language Functions: interpret, inquire	
Language Use Context: social studies; school navigational/instructional	
Language Modality Integrated: listen & read & write	
Pragmatic Arrangements Communicative mode: interpretive, productive Speaker: adult- school principal Audience: adult- visiting journalist	
Genre: informational	
Language Processing Skill: basic	<div style="background-color: #ADD8E6; padding: 5px;"> 3 GUEST Now write the first question for the visitor here. </div>

Note: The ELP item is a released sample writing task from the English Language Proficiency Assessment for the 21st Century. Copyright 2021 by ELP21. Reprinted with permission.

We have presented one way to comprehensively analyze the language demands in standards and assessments based on our protocol and its application. The level of analysis details will certainly depend on the purpose of alignment evaluation. One may choose a group of specific language complexity categories to evaluate the alignment of language demands for practicality and efficiency.

Fourth, in applying language- and/or cognitive-complexity protocols for ELP assessments, qualifications of coders and quality of coding decisions are of crucial importance. In our project, agreement on characterizing language demands between trained coders with linguistics expertise and prior

English as a second language (ESL) or English as a foreign language teaching experience was variable; agreement also differed depending on specific coding category. Lin and Zhang’s (2014) review of 20 standards-correspondence studies that used content experts and ESL specialists to rate cognitive complexity in the language performance indicators showed that “depending on the particular grades and subject areas, 3–6 reviewers are needed to achieve acceptable reliability and to control for reasonable measurement errors in their judgments” (p. 413; see also Christopherson & Webb, 2015, which discussed the challenge and complexity of rater agreement in this domain). Focus groups with teachers who were part of the William T. Grant Foundation project

revealed how general education teachers and ESL specialists also struggled with the ambiguity of the standards descriptions, which can lead to different understandings of the language demands of the standards (Wolf et al., 2021).

We hope that these findings and insights on implementing protocols and procedures to characterize language demands in a principled way can assist states and others to make meaning of language constructs across the standards and inform their generation of evidence to argue for alignment between ELP standards and academic content standards and alignment between ELP standards and ELP assessments. In the next section, we make further suggestions to consider when collecting evidence for ELP alignment.

Objective 3: Key considerations for ELP alignment evaluation

In an effort to examine the language demands in ELP standards and academic content standards, Christopherson and Webb (2015) utilized the key practice language functions described in the *Framework for English Language Proficiency Development Standards Corresponding to the Common Core State Standards and the Next Generation Science Standards* (CCSSO, 2012). Christopherson and Webb performed standards-to-standards alignment with the model performance indicators (MPIs) of the 2007 and 2012 editions of the WIDA English language proficiency/development standards (considered equivalent artifacts with their focus on language features). Additionally, they reported the results of a correspondence study between the Florida calculus standards and the WIDA MPIs. Overall, they reported relatively weak

correspondences. While Christopherson and Webb's work represents a post-ESSA approach to examining language demands by adopting new methods, innovations are still needed that can effectively capture the "English language knowledge and skills necessary to reflect the language needed to acquire and demonstrate the skills identified in the State's academic content standards" (USDOE, 2018, p. 31). In a similar vein, WestEd (2015) conducted a correspondence study to examine the alignment of California's English language development standards to academic content standards. In this study, researchers used the mathematical key practices and the science/engineering key practices delineated in the Common Core State Standards for Mathematics and the Next Generation Science Standards, respectively. By using a small set of high-level key practice descriptions instead of each individual standard description, the correspondence evaluation was performed at the macro level, yet little information was reported about rater agreement in the WestEd study. Building on previous work, we strove to deepen and refine our protocols, streamlining the characterizations of language and increasing reliability across features, as well as deploying them with additional relevant artifacts such as test items and curricular materials.

It is evident that more empirical research is needed to shed light on ELP alignment methods. In the interim, we recommend SEAs, assessment developers, and/or alignment evaluators lay out the level of alignment details and the focus of alignment evaluation in determining the alignment method, including the categories of examination. We offer the following foundational questions for use by SEAs to ascertain the quality of the evidence generated by the required ELP alignment evaluation. Such questions may also

play a role in drafting requests for proposals for alignment studies and vetting responses from vendors/consultants who may be contracted to conduct such studies.

1. How are language demands operationalized? Will the construct cover language knowledge, skills, and language use across disciplines? How are language skills/modalities presented in standards?
2. Once the constructs of the ELP assessments are defined (based on the analysis of language demands in ELP and content-area standards), which standards and language demands are represented in ELP assessments? To what extent are each standard and language demand covered in ELP assessments? (This information may be documented in the test specifications or blueprint.)
3. What aspects of language demand/complexity are to be examined for alignment (e.g., range of language functions; language content coverage related to content areas; pragmatic arrangement features, such as speaking to the whole group, conversing in dyads)?
4. What approaches to alignment evaluation are being proposed? How will traditional alignment approaches used for academic content standards and assessments, such as the Depth of Knowledge scale (Webb, 2005), be modified to characterize language complexity?
5. What unit of the standards (or test items) is useful for judging alignment? How are the different language modalities (i.e., listening, reading, speaking, writing) treated in alignment (e.g., discrete, integrated)?
6. What criteria will be implemented in selecting the coders who characterize the language demands? How and how well will coders be trained?

7. How will reliability between coders be calculated? What levels of reliability can be expected? What process will be implemented in resolving coder disagreements?
8. How will an acceptable level of alignment be determined? How will the information provided by protocol coders be compiled and used to identify areas of acceptable coverage as well as any gaps or weaknesses (e.g., new assessment item development, standards review and revision)?

Remaining challenges and opportunities

Inclusion of ELP assessments in peer review is relatively new compared to the long-standing tradition of peer review on academic content assessments (Farnsworth, 2020). While the assessment peer review process (USDOE, 2018) has raised the significance of ELP alignment evaluation substantially, the guidance was made based on a limited body of ELP alignment methods. The questions raised in the previous section beg for more empirical research and discussion among relevant stakeholders. For instance, which categories of our language demand protocol are most practical and parsimonious for ELP assessment alignment evaluation? Take the “acceptable level of alignment” as an example. Once the degree of match between different standards or between standards and assessment items is determined, there is still the issue of how to interpret these data. For example, alignment expectations have been set at 75% for the match between compared units (e.g., Christopherson & Webb, 2015). But are these percentages of alignment sufficient for peer review?

More collective discussion based on empirical research (e.g., understanding of student outcomes tied to different degrees of alignment) is needed to address this question.

Establishing ELP alignment is a continuous improvement process that involves the study of the language demands embedded in the content of standards and assessments. Additionally, ELP alignment should also be concerned with the language complexity presented in curricular materials as well as how EL students are performing on assessments with various types of language complexity. The categories to characterize the language complexity employed in the current research project may also be applied to curricular materials to ensure alignment among standards, assessments, and curricular/instructional materials. This line of study will provide technically sound methods to evaluate the alignment of ELP assessments and their related standards in a systematic and sustainable manner. Further, they will in turn help the assessment peer review process continuously improve the evidence-based guidelines. Above all, continuous ELP alignment evaluation will help to ensure the equity of supporting EL students in standards-based education.

References

- Bailey, A. L. (2015). *Pragmatic arrangements of ELP/D standards and ELP assessment* [Unpublished manuscript]. University of California, Los Angeles.
- Bailey, A. L., Butler, F. A., & Sato, E. (2007). Standards-to-standards linkage under Title III: Exploring common language demands in ELD and science standards. *Applied Measurement in Education*, 20(1), 53–78. <https://doi.org/10.1080/08957340709336730>
- Chi, Y., Lin, C.-K., & Zhang, J. (2011). *Alignment between the WIDA English Language Proficiency Standards, 2007 edition, pre-kindergarten through Grade 12 and ACCESS for ELLs assessment*. Wisconsin Center for Education Research.
- Christopherson, S. C., & Webb, N. L. (2015). *Final report: Alignment analysis of key practice language functions from the framework for English Language Proficiency Development Standards corresponding to the Common Core State Standards for English Language Arts and Mathematics and the WIDA English Language Proficiency Standards*. Wisconsin Center for Education Products & Services.
- Council of Chief State School Officers. (2012). *Framework for English language proficiency development standards corresponding to the Common Core State Standards and the Next Generation Science Standards*. <https://www.ccsso.org/sites/default/files/2017-11/ELPD%20Framework%20Booklet-Final%20for%20web.pdf>
- Council of Chief State School Officers. (2014). *English language proficiency (ELP) standards with correspondences to K-12 English language arts (ELA), mathematics, and science practices, K-12 ELA standards, and 6-12 literacy standards*. https://ccsso.org/sites/default/files/2017-11/Final%204_30%20ELPA21%20Standards%281%29.pdf
- Every Student Succeeds Act, 20 U.S.C. § 6301 (2015).
- Farnsworth, T. (2020). A review of validity evidence on K-12 English language proficiency assessment. In M. K. Wolf (Ed.), *Assessing English language proficiency in U.S. K-12 schools* (pp. 75–91). Routledge. <https://doi.org/10.4324/9780429491689-5>
- Lin, C.-K., & Zhang, J. (2014). Investigating correspondence between language proficiency standards and academic content standards: A generalizability theory study. *Language Testing*, 31(4), 413–431. <https://doi.org/10.1177/0265532213520304>
- Paradis, J., Rusk, B., Duncan, T. S., & Govindarajan, K. (2017). Children's second language acquisition of English complex syntax: The role of age, input, and cognitive factors. *Annual Review of Applied Linguistics*, 37, 148–167. <https://doi.org/10.1017/S0267190517000022>
- U.S. Department of Education. (2018). *A state's guide to the U.S. Department of Education's assessment peer review process*. <https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>
- Webb, N. L. (2005). *Web alignment tool (WAT) training manual*. Wisconsin Center for Education Research.
- WestEd. (2015). *Correspondence study report: California English Language Development Standards, Common Core State Standards for Mathematics, and Next Generation Science Standards: Correspondence study and augmentation document project*. <https://www.cde.ca.gov/ta/tg/ep/documents/correspond-yrptab899.pdf>
- WIDA. (2014). *2012 amplification of the WIDA English language development standards, kindergarten–Grade 12*. <https://wida.wisc.edu/sites/default/files/resource/2012-ELD-Standards.pdf>
- WIDA. (2020). *WIDA English language development standards framework, 2020 Edition, kindergarten—Grade 12*. <https://wida.wisc.edu/sites/default/files/resource/WIDA-ELD-Standards-Framework-2020.pdf>
- Wolf, M. K., Bailey, A. L., Ballard, L., Wang, Y., & Pogossian, A. (2021, April 8–12). *Understanding the language demands in academic content and English language proficiency standards for English learners* [Paper presentation]. The virtual annual meeting of the American Educational Research Association.

Appendix: Selected literature relevant to language demand protocol development

Note: This is not a comprehensive list of literature.

- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford University Press.
- Bailey, A. L., Butler, F. A., Stevens, R., & Lord, C. (2007). Further specifying the language demands of school. In A. L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 103–156). Yale University Press.
- Bailey, A. L., & Heritage, M. (2019). *Progressing students' language day by day*. Sage/Corwin Press.
- Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Quarterly*, 48(3), 48–506. <https://doi.org/10.1002/tesq.176>
- Bailey, A., & Wolf, M. K. (2020). The construct of English language proficiency in consideration of college and career readiness standards. In M. K. Wolf (Ed.), *Assessing English language proficiency in U.S. K-12 schools* (pp. 36–54). Routledge. <https://doi.org/10.4324/9780429491689-3>
- Bruner, J. (1990). *Acts of meaning*. Harvard University Press.
- Butler, F. A., Bailey, A. L., Stevens, R., Huang, B., & Lord, C. (2004). *Academic English in fifth-grade mathematics, science, and social studies textbooks* (CSE Technical Report No. 642). University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Christie, F. (2012). *Language education throughout the school years: A functional perspective*. Wiley-Blackwell.
- Cook, H. G. (2005). *Aligning English language proficiency tests to English language learning standards*. Council of Chief State School Officers.
- Cook, H. G., & MacDonald, R. (2013). *Tool to evaluate language complexity of test items* (WCER Working Paper No. 2013-5). University of Wisconsin-Madison, Wisconsin Center for Education Research.
- Council of Chief State School Officers. (2012). *Framework for English Language Proficiency Development Standards corresponding to the Common Core State Standards and the Next Generation Science Standards*.
- Gebhard, M., & Harman, R. (2011). Genre theory in K-12 schools: A response to high-stakes school reforms in the United States. *Journal of Second Language Writing*, 20(1), 45–55. <https://doi.org/10.1016/j.jslw.2010.12.007>
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Halliday, M. A. K., & Hasan, R. (1989). *Language, context, text: Aspects of language in a social-semiotic perspective*. Oxford University Press.
- Levinson, S. C. (1980). Speech act theory: The state of the art. *Language Teaching*, 13(1–2), 5–24. <https://doi.org/10.1017/S0261444800008661>
- Sato, E. (2010). *Language for achievement: A framework for academic English language*. WestEd.
- Scarcella, R. (2003). *Academic English: A conceptual framework* (Technical

Report No. 2003-1). University of California, Linguistic Minority Research Institute.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Lawrence Erlbaum. <https://doi.org/10.4324/9781410610317>

Wolf, M. K., Herman, J. L., Kim, J., Abedi, J., Leon, S., Griffin, N., Bachman, P. L., Chang, S. M., Farnsworth, T., Jung, H., Nollner, J., & Shin, H. W. (2008). *Providing validity evidence to improve the assessment of English language learners* (CREST Technical Report No. 738). University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Notes

- 1 While the project focus is on standards, assessments, and curricular material at the fifth grade using the Common Core State Standards, WIDA, and the English language standards of the English Language Proficiency Assessment for the 21st Century (ELPA21), the procedures and many findings will be relevant for other grades and other states that have adopted identical or equivalent ELP and academic content standards and their linked assessments. The operationalization of language demands in order to characterize linguistic complexity at fifth grade can be extended to other grades given the ELP standards are formulated in comparable ways across grades and are largely representative of progressively sophisticated examples of many similar language constructs spanning kindergarten through 12th grade.
- 2 We analyzed the 2012 Amplification of WIDA English Language Development Standards (WIDA, 2014), which was written to specifically address Common Core State Standards correspondence to WIDA standards. The more recent 2020 edition of the WIDA standards was not yet released at the start of the project.

Acknowledgments

This work was supported by funding from the William T. Grant Foundation grant (ID 187863) to Educational Testing Service (ETS) and University of California, Los Angeles (UCLA). The opinions expressed are those of the authors and do not represent views of the William T. Grant Foundation, ETS, or UCLA. The authors would like to thank David Anderson, Melissa Gholson, John Norris, and Jonathan Schmidgall at ETS and Anahit Pogossian at UCLA for their helpful comments on an earlier draft of this paper. The authors also thank Kim Fryer for her valuable assistance in copy-editing and publishing this paper.

Alison L. Bailey is a professor and division head of Human Development and Psychology in the Department of Education, UCLA; abailey@gseis.ucla.edu.

Mikyung Kim Wolf is a managing principal research scientist in the Center for Language Education and Assessment Research, ETS; mkwolf@ets.org.

Laura Ballard is a senior research project manager in the Center for Language Education and Assessment Research, ETS; lballard@ets.org.

Suggested citation: Bailey, A. L., Wolf, M. K., & Ballard, L. (2022). *English language proficiency standards and assessment alignment: Issues, insights, and innovations for guiding peer review* (ETS Research Note). ETS.

ETS grants users permission to copy and redistribute the material in any medium or format for noncommercial purposes. Each reproduction of any part of the materials must contain notice of Educational Testing Service's copyright as follows:

Copyright © 2022 by Educational Testing Service. All rights reserved. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Photographs and figures in this publication are not to be used in the promotion of any commercial product or in other publications without further permission from the respective copyright owners.

ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.